H. M. IONESCU

Elemente de statistică matematică

EDITURA ȘTIINȚIFICĂ

HARALAMBIE M. IONESCU

M. Galoteanu

ELEMENTE DE STATISTICĂ MATEMATICĂ

EDITURA ȘTIINȚIFICĂ
București — 1957

Scanned with OKEN Scanner

DIN PARTEA AUTORULUI

În această lucrare sînt expuse principiile teoretice de bază ale statisticii matematice.

Pentru înțelegerea cursului de statistică matematică, cunoștințele de matematici strict necesare unui statistician sau economist sînt suficiente. Tocmai din aceste motive o parte din problemele care reclamă un aparat matematic mai complicat au fost omise, iar pentru altele s-au dat numai unele indicații utile pentru studii mai aprofundate. Teoriile expuse sînt, pe cît posibil, ilustrate prin exemple simple. De asemenea, cursul tratează majoritatea problemelor pe care trebuie să le conțină un curs modern de statistică matematică. Tot pentru o mai simplă expunere în acest manual nu sînt tratate problemele neparametrice și nici metodele de verificare a ipotezelor statistice care trebuie să-și găsească în mod necesar locul într-un curs complet.

Un tratat de statistică matematică n-a existat pînă în momentul de față în literatura noastră de specialitate și de aceea cred că actualul manual va avea lipsuri pe care numai experiența urmează să le corecteze.

La alcătuirea lui au fost folosite tratatele străine, îndeosebi cele sovietice; o parte din materiale mi le-a pus la dispoziție prof. Gh. Mihoc, decanul Facultății de Matematică și Fizică din București, căruia îi exprim viile mele mulțumiri.

Cursul se adresează îndeosebi studenților institutelor cu caracter economic și tehnic și tehnicienilor din întreprinderile care aplică în mod practic metodele statisticii matematice.

H. M. IONESCU

Scanned with OKEN Scanner

INTRODUCERE

Scurt istoric al dezvoltării statisticii matematice

Teoria erorilor elaborată de Laplace (1749—1827), Legendre şi Gauss este considerată ca prima lucrare de statistică matematică. Adolphe Quételet (1796—1874) aplică teoria probabilităţilor la studiul fenomenelor sociale. Lucrările lui Francis Galton (1822—1907) sînt considerate începutul statisticii matematice moderne. El a pus bazele teoriei corelaţiei şi a aplicat în mod sistematic metodele statisticii matematice în biometrie, al cărei întemeietor este de altfel. Ca reprezentanţi ai statisticii matematice moderne citäm pe K. Pearson, continuatorul ideilor lui Galton în teoria corelaţiei; el a elaborat un studiu aprofundat al distribuţiilor curbe şi a adus importante contribuţii în teoria selecţiei. R. A. Fisher are contribuţii de seamă în teoria selecţiilor mici, a creat o nouă metodă a analizei dispersionale şi a dezvoltat teoria verosimilităţii maxime, care joacă un rol de bază în teoria estimaţiilor.

Paralel cu dezvoltarea şcolii statisticii matematice engleze, în a doua jumătate a secolului trecut au apărut în Rusia lucrările fundamentale ale lui A.A. Ciuprov asupra distribuţiilor statistice, asupra stabilităţii frecvenţelor şi corelaţiilor. Lucrările marilor probabilişti ruşi P. L. Cebîşev, A. Markov şi Leapunov au avut o importanţă considerabilă pentru dezvoltarea statisticii matematice. În lucrările savanţilor sovietici sînt rezolvate o serie de probleme fundamentale pentru statistica matematică. Menţionăm îndeosebi lucrările lui S. N. Bernstein, Glivenko, V. Romanovski, A.N. Kolmogorov, A.I. Hincin şi B.V. Gnedenko.

Sub denumirea de statistică matematică se înţelegea o disciplină specială care fundamenta diferite metode de studiu ale fenomenelor biologice de natalitate şi ereditate. Acest gen de probleme forma obiectivul principal al cercetărilor şcolii

Scanned with OKEN Scanner

engleze. Viața și în primul rînd dezvoltarea tehnicii au făcut ca statistica matematică să se dezvolte într-un ritm din ce în ce mai viu, să se recunoască caracterul său științific și să se elaboreze riguros bazele sale științifice. În această privință un mare rol a avut școala statisticii matematice sovietice. În concepția actuală, prin statistică matematică se înțelege *elaborarea metodelor care permit să se obțină din datele experienței concluzii fundamentate din punct de vedere științific asupra fenomenelor de masă în care se manifestă întîmplarea, fenomene în care același complex de condiții se repetă de mai multe ori.*

Înțelegem prin *fenomen de masă, un fenomen care se compune dintr-un mare număr de fenomene individuale care se produc în urma realizării în practică a unui anumit complex de condiții bine determinate.*

Statistica matematică se sprijină pe teoria probabilităților și studiază legile generale ale fenomenelor de masă într-o formă abstractă, fără a ține seama de natura particulară a obiectelor cercetate. Din această cauză concluziile statisticii matematice se aplică la fenomene foarte variate din diferite domenii : biologie, fizică, chimie, zootehnie, agricultură, precum și în problemele de control și încercare ale produselor industriale etc. Trebuie însă să atragem atenția asupra faptului că valabilitatea aplicării metodelor statisticii matematice nu poate fi apreciată cu ajutorul unor criterii matematice arbitrare, formale. Specificul cercetării statistice a unui fenomen de masă constă în aceea că, de obicei, atenția este concentrată asupra unor caracteristici ale colectivității pe care o cercetăm, caracteristici pe care le socotim fundamentale. De aceea, în analiza și compararea proceselor de masă apar probleme concrete care impun metode specifice de studiu și o analiză profundă a legăturilor lor cauzale, așa încît analiza calitativă înaintea celei cantitative este un principiu de bază în aplicarea statisticii matematice la problemele concrete.

★

Materia tratată în cursul de față a fost împărțită în 11 capitole, în mod gradat. În primul capitol sînt prezentate sumar cunoștințele strict necesare de statistică teoretică expuse într-o formă care permite să facem legătura dintre noțiunile bazate pe datele experienței și noțiunile abstracte ale teoriei probabilităților pe baza căreia este construită întreaga statistică matematică.

Al doilea capitol tratează cîmpul de evenimente aleatorii și noțiunile de bază cu privire la probele independente între

Scanned with OKEN Scanner

ele (probele stohastice) şi se lămureşte noţiunea fundamentală de evenimente elementare.

Al treilea capitol conţine noţiunile strict necesare din teoria probabilităţilor pe care se bazează teoremele statisticii matematice.

În capitolele IV şi V sînt tratate distribuţiile statistice, se definesc noţiunile de variabilă aleatorie de tip discret şi continuu, densitatea de repartiţie şi se descriu distribuţiile statistice discrete şi continue care se întîlnesc în mod frecvent în aplicaţii.

Capitolul VI tratează despre valorile probabile (sau valorile medii) şi despre momente. Aci se introduc noţiunile importante de funcţie generatoare şi funcţie caracteristică.

Capitolul VII conţine distribuţiile funcţiilor de variabile aleatorii şi se aprofundează unele distribuţii des întîlnite în practică : distribuţia normală cu mai multe dimensiuni şi problema regresiunii, distribuţia χ^2, distribuţia lui R. A. Fisher şi distribuţia lui Student.

În capitolul VIII se introduc noţiunile de convergenţă în probabilitate şi se expun teoremele lui Cebîşev, Laplace, Leapunov şi Bernoulli, teoreme care în ansamblul lor reprezintă principiul general al legii numerelor mari.

Capitolul IX conţine noţiunile fundamentale asupra selecţiilor aleatorii şi se demonstrează teorema centrală de limită în virtutea căreia media de selecţie are o distribuţie asimptotic normală, oricare ar fi distribuţia populaţiei din care s-a extras selecţia. Un paragraf important este cel al selecţiilor mici care sînt folosite pe scară largă în producţia industrială.

În capitolul X sînt tratate problemele parametrice şi estimaţiile parametrilor distribuţiilor specificate. Se definesc noţiunile de estimaţii eficiente şi suficiente, iar ca metode de calcul se expune metoda momentelor şi metoda verosimilităţii maxime. Un paragraf al acestui capitol a fost consacrat estimaţiilor pentru selecţiile nerepetate care sînt curent utilizate în practica statistică.

Ultimul capitol tratează problema estimării intervalelor şi regiunilor de confidenţă pentru parametrii distribuţiilor statistice. Se expun două metode elementare în acest scop şi se fac aplicaţii pentru determinarea parametrilor distribuţiei normale cu o singură dimensiune.

La sfîrşitul cursului se dau patru tabele statistice, utile pentru efectuarea calculelor şi verificarea exemplelor din cuprinsul cărţii.

Scanned with OKEN Scanner

Scanned with OKEN Scanner

Capitolul I

NOȚIUNI INTRODUCTIVE

§ 1. Problema selecției

Sarcina statisticii matematice constă în elaborarea metodelor care să permită descoperirea trăsăturilor fundamentale ale fenomenului sau procesului de masă cercetat. În acest scop se folosește un material, de obicei restrîns, obținut pe cale experimentală, adică prin efectuarea unor probe în general independente între ele. Desigur că acest material prezintă în primul rînd un interes practic imediat, deoarece el ne furnizează informații asupra fenomenului sau procesului observat și ca atare pe baza lui putem caracteriza un anumit proces concret. Dacă de pildă dispunem de un material statistic în legătură cu piesele fabricate de o anumită mașină în anumite condiții, se înțelege că determinarea numărului de piese defecte și bune este o operație utilă și ea ne dă o primă imagine asupra procesului de fabricație cercetat. Problemele statisticii matematice apar însă atunci cînd pe baza materialului obținut prin experiență trebuie să caracterizăm fenomenul în ansamblul său, adică să determinăm probabilitatea ca o piesă fabricată să fie o piesă rebut, valoarea medie a numărului acestor piese, distribuția lor etc.

Așadar, statistica matematică studiază cu ajutorul teoriei probabilităților *colectivitățile statistice* sau *populațiile* după denumirea curentă. (Piesele fabricate de o mașină constituie, de exemplu o populație, prețurile produselor alimentare constituie o populație, mulțimea stemelor apărute aruncînd o monedă este de asemenea o populație etc.).

Revenind la populația formată din piesele fabricate de o anumită mașină, în condiții normale de funcționare a mașinii,

piesele standard şi cele rebut se distribuie după o lege de repartiţie cunoscută. Cunoaşterea acestei distribuţii este suficientă pentru a calcula numărul pieselor standard fabricate într-un interval de timp dat, dispersia acestora de la media lor, precum şi alte caracteristici probabilistice. De asemenea vom putea constata dacă maşina funcţionează în condiţii normale sau nu.

Distribuţia de care vorbeam mai sus poate fi determinată printr-o observare totală, adică cercetînd toate elementele populaţiei; în unele împrejurări însă este nevoie să cunoaştem numai anumite caracteristici ale populaţiei şi dacă prin anumite consideraţii de ordin teoretic nu putem determina distribuţia acestei populaţii este posibil ca în baza unui număr oarecare de rezultate ale unor probe succesive să calculăm caracteristicile care ne interesează.

Metoda folosită în acest scop este cunoscută sub numele de „*metoda selectivă*" şi ea constituie unul din cele mai importante capitole ale statisticii matematice. Într-un mod simplu, metoda selecţiei o putem enunţa astfel: «se dă populaţia Γ formată din N elemente. Efectuînd n probe succesive obţinem rezultatele

$$x_1, x_2 \ldots x_n \qquad (1)$$

Pe baza acestor rezultate să se determine caracteristicile populaţiei Γ (media, mediana, momentele de diferite ordine etc.)».

Dacă numărul N este infinit, singura metodă indicată pentru determinarea diferitelor caracteristici ale populaţiei este metoda selectivă. De exemplu, în medicină, pentru determinarea numărului mediu de corpuscule aflate în sînge, se ia o cantitate relativ mică de sînge şi rezultatul analizei acesteia se extinde asupra întregii cantităţi de sînge din organism. Asemenea analize se fac, după cum se ştie, în toate ştiinţele experimentale. Metoda selectivă însă nu foloseşte numai pentru cercetarea populaţiilor infinite; ea este utilizată şi cînd N este un număr finit şi în unele împrejurări constituie singura metodă ştiinţifică de cercetare. Aşa de pildă ea se utilizează în cazul cînd cercetarea tuturor elementelor populaţiei ar duce la distrugerea acestor elemente, cum este cazul în industria textilă cînd se stabileşte rezistenţa unui anumit produs, în industria chimică, în industria fabricării becurilor electrice etc. Metoda selectivă se întrebuinţează, de asemenea, în practica recensămintelor; ea prezintă, după cum se ştie, o serie de avantaje faţă de observarea totală, cu analiza cărora se ocupă îndeajuns statistica teoretică.

Scanned with OKEN Scanner

Este clar că statistica matematică nu se ocupă decît de selecțiile făcute la întîmplare, *înțelegînd prin selecții, grupul de elemente din populație, observat prin probele repetate.* Numărul elementelor din grup se numește *volumul selecției.* Așa de exemplu selecția (1) este o *selecție de volum n*[1]. Practic, selecția se efectuează prin diferite procedee; în particular, vom descrie două dintre procedeele fundamentale în metoda selectivă și anume:

1°. *Selecția repetată*, în care elementul observat este înapoiat în colectivitate așa încît la proba următoare el poate fi cercetat din nou.

2°. *Selecția nerepetată*, în care elementul cercetat nu se mai înapoiază în colectivitate.

Evident că pentru populațiile infinite problema selecției cu sau fără înapoiere nu se mai pune. Am precizat mai sus că statistica matematică nu se ocupă decît cu selecțiile întîmplătoare și aceasta pentru motivul foarte important că o selecție întîmplătoare face posibilă calcularea unei limite a erorii comise, în timp ce o selecție care nu are un caracter întîmplător nu poate da nici un fel de măsură a gradului de siguranță care poate fi atribuit rezultatelor obținute din selecție.

§ 2. Serii statistice. Reprezentări geometrice

Să considerăm tabloul de mai jos în care este redată repartiția unui număr de 11 450 familii, după numărul copiilor în viață:

Tabloul 1

Numărul copiilor în viață	Numărul familiilor
0	1 892
1	3 689
2	2 798
3	1 550
4	831
5	455
6	235
	11 450

Acest tablou reprezintă o selecție de volum n = 11 450, dintr-o populație Γ. Dacă pe această populație definim o variabilă x, să zicem numărul copiilor în viață, selecția de mai sus ne conduce la cele două șiruri de valori din tabloul 1: de o parte mulțimea valorilor lui x (coloana I), iar pe de altă parte, și în corespondență cu acestea numărul acelor elemente extrase din Γ pentru care x = 0, 1, 2, ...6 (coloana II). Notînd cu n(x) numărul elementelor pentru care variabila x ia valoarea x,

[1] Selecției (1) i se mai spune și „eșantion" și se însemnează de obicei $[x_1, x_2, \ldots x_n]$.

adică $n(0) = 1\,892$, $n(1) = 3\,689$, ...$n(6) = 235$ avem evident :

$$\sum_{x=0}^{6} n(x) = 11\,450$$

Dacă variabila ia valorile 0, 1, 2, ..., r avem relația :

$$\sum_{x=0}^{r} n(x) = n \tag{2}$$

Mai general, dacă variabila x definită pe mulțimea Γ ia valorile $x_1, x_2, \ldots, x_r$ și dacă $n(x_k)$ reprezintă numărul elementelor din selecția de volum r, pentru care $x = x_k$, $1 \leqslant k \leqslant r$, avem :

$$\sum_{k=1}^{r} n(x_k) = n \tag{2_1}$$

Numerele $n(x_k)$ se numesc *frecvențele absolute ale valorilor* x_k $(1 \leqslant k \leqslant r)$ ale variabilei x.

Tabloul 1 care reprezintă corespondența dintre valorile variabilei și frecvențele acestor valori este numit *seria statistică* sau *distribuția variabilei x.*

Prin definiție, raportul

$$f(x_k) = \frac{n(x_k)}{n}, \quad 1 \leqslant k \leqslant r \tag{2_2}$$

se numește *frecvența relativă a valorii* x_k.

Deoarece avem din (2_1) :

$$\sum_{k=1}^{r} n(x_k) = n, \text{ atunci } \quad \frac{\sum_{k=1}^{r} n(x_k)}{n} = \sum_{k=1}^{r} \frac{n(x_k)}{n} = 1$$

rezultă că suma frecvențelor relative este egală cu 1 :

$$\sum_{k=1}^{r} f(x_k) = 1 \tag{2_3}$$

Scanned with OKEN Scanner

În exemplul considerat avem frecvențele din tabloul 2 :

Tabloul 2

x_k	$n(x_k)$	$f(x_k)$
0	1 892	0,165
1	3 689	0,322
2	2 798	0,244
3	1 550	0,135
4	831	0,073
5	455	0,040
6	235	0,021
	11 450	1,—

Observație. De obicei, în aplicații se folosesc, în locul frecvențelor relative $f(x_k)$, procentele valorilor x_k, adică

$$\theta_k = 100\ f(x_k) \qquad (2_4)$$

Avem evident relația :

$$\sum_{k=1}^{r} \theta_k = 100 \qquad (2_5)$$

Din tabloul 2 deducem procentele θ_k

Tabloul 3

x_k	θ_k
0	16,5
1	32,2
2	24,4
3	13,5
4	7,3
5	4,0
6	2,1
	100,—

Distribuția din tabloul 1 se poate reprezenta geometric însemnînd pe axa absciselor valorile 0, 1, 2, 3, 4, 5, 6, iar pe ordonatele respective fie frecvențele $n(x_k)$, fie $f(x_k)$, fie procentele θ_k.

Pentru scopurile statisticii matematice, noi vom folosi frecvențele relative $f(x_k)$.

Graficul are înfățișarea de mai jos.

În statistica aplicată, pentru mai multă expresivitate, ordonatele simple din grafic se înlocuiesc prin dreptunghiuri cu bazele egale, de lungime arbitrară.

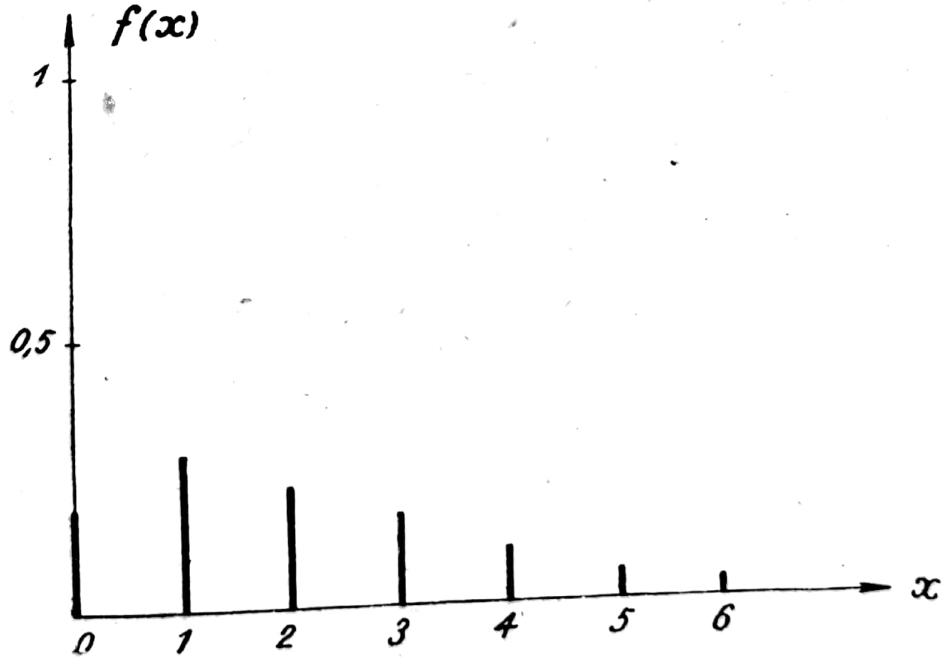

Fig. 1

Scanned with OKEN Scanner

Funcția de repartiție empirică. Vom numi *funcție de repartiție empirică a variabilei* x, sau *funcție de repartiție a selecției de volum* n funcția

$$F_n(x) = \sum_{x_k \leqslant x} f(x_k) \tag{3}$$

suma extinzîndu-se asupra tuturor valorilor variabilei, mai mici decît o valoare dată x.

Dacă valorile variabilei sînt așezate în ordinea mărimii crescătoare, adică x_1 este cea mai mică valoare și x_r cea mai mare, avem tabloul:

$$F_n(x) = \begin{cases} 0, & \text{pentru } x \leqslant x_1 \\ f(x_1) & ,, \quad x_1 \leqslant x < x_2 \\ f(x_1) + f(x_2) = \sum_{k=1}^{2} f(x_k) & ,, \quad x_2 \leqslant x < x_3 \\ \cdots & \\ \sum_{k=1}^{r-1} f(x_k) & ,, \quad x_{r-1} \leqslant x < x_r \\ \sum_{k=1}^{r} f(x_k) = 1 & ,, \quad x_r \leqslant x \end{cases}$$

Funcția de repartiție pentru tabloul 1 are valorile:

Tabloul 4

x_k	$f(x_k)$	$F_n(x_k)$
0	0,165	0,165
1	0,322	0,487
2	0,244	0,731
3	0,135	0,866
4	0,073	0,939
5	0,040	0,979
6	0,021	1,—
	1,—	

Scanned with OKEN Scanner

Graficul acestei funcţii este următorul :

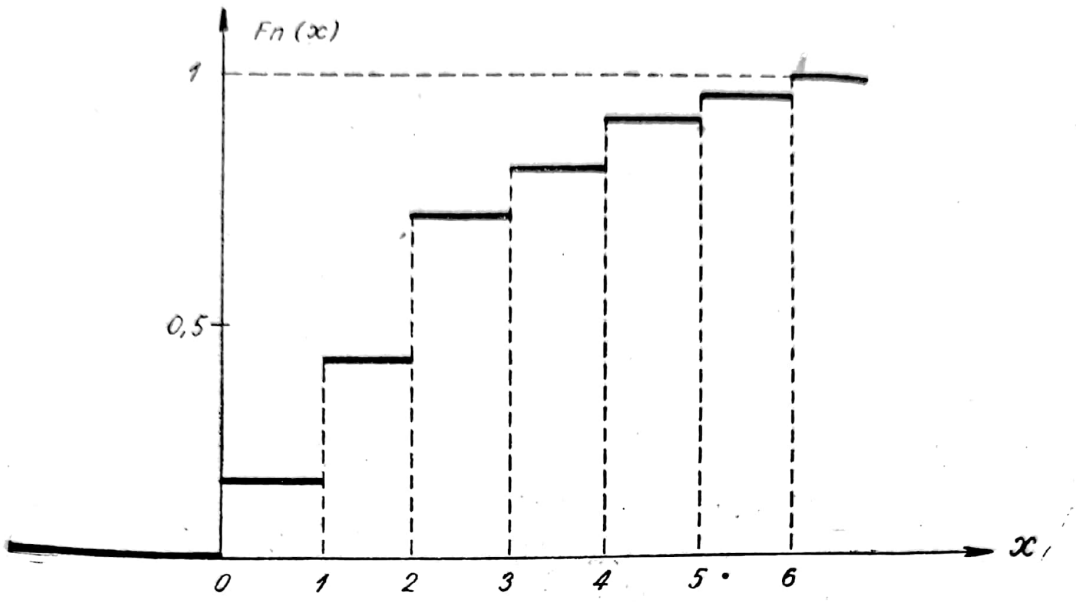

Fig. 2

Funcţiile f (x) sau F_n (x) determină complet distribuţia[1] variabilei x. În adevăr, avînd valorile f (x_k) şi volumul n al selecţiei şi înlocuind în relaţia (2_2) :

$$n(x_k) = n f(x_k), \qquad 1 \leqslant k \leqslant r, \tag{$2'_2$}$$

înseamnă că pentru orice valoare a variabilei x obţinem şi frecvenţele absolute n (x).

Dacă dimpotrivă se dau valorile funcţiei de repartiţie, din (3) deducem :

$$F_n(x_{k+1}) = f(x_1) + f(x_2) + \ldots + f(x_{k-1}) + f(x_k)$$
$$F_n(x_k) = f(x_1) + f(x_2) + \ldots + f(x_{k-1})$$

De unde, prin scăderea celor două relaţii obţinem frecvenţele relative

$$f(x_k) = F_n(x_{k+1}) - F_n(x_k), \tag{3_1}$$

iar din ($2'_2$) obţinem frecvenţele absolute.

Observaţie. Dacă valorile $x_1, x_2, \ldots, x_k$, obţinute prin selecţie, sînt distincte între ele, atunci

$$n(x_k) = 1, \qquad f(x_k) = \frac{1}{n} \tag{3_2}$$

pentru orice $k = 1, 2, \ldots, n$

[1] Prin distribuţia variabilei x înţelegem valorile n (x) corespunzătoare valorii x. Această noţiune va fi reluată mai tîrziu sub alt aspect.

Funcția de repartiție empirică va fi egală cu raportul

$$F_n(x) = \frac{n_x}{n} \qquad (3')$$

unde n_x reprezintă numărul valorilor variabilei mai mici decît x dat.

Exemplu. Să presupunem că din colectivitatea Γ considerată în exemplul precedent s-a făcut o selecție de volum n = 5, obținîndu-se rezultatele :

Tabloul 5

Numărul copiilor în viață :	0,	1,	2,	3,	4,	5, 6
Numărul familiilor :	1,	1,	1,	0,	1,	0, 1

Pentru o atare distribuție avem următorul calcul :

Tabloul 6

x_k	$n(x_k)$	$f(x_k)$	$F_n(x_k)$
0	1	0,2	0,2
1	1	0,2	0,4
2	1	0,2	0,6
3	0	0	0,6
4	1	0,2	0,8
5	0	0	0,8
6	1	0,2	1,—
	5	1,—	

Graficele frecvențelor și ale funcției de repartiție sînt cele redate mai jos :

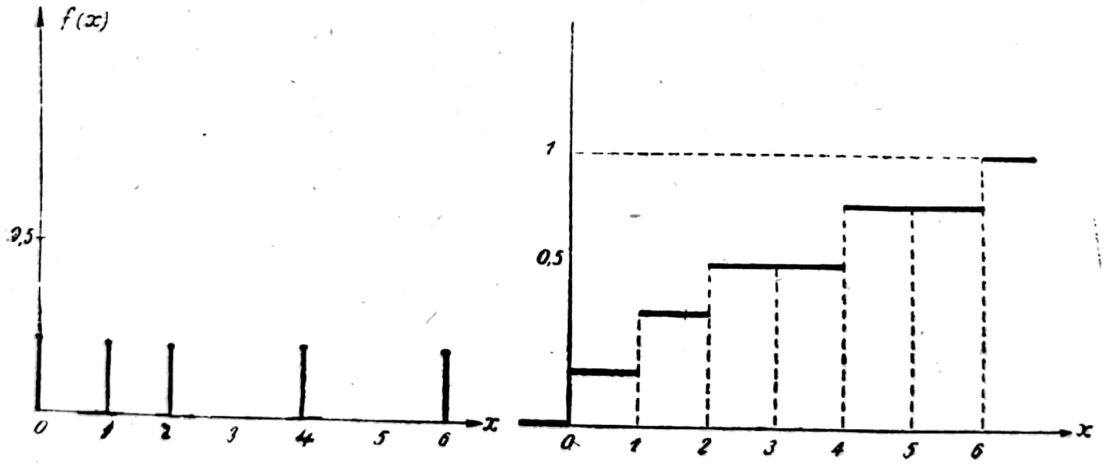

Fig. 3 și 4

Variabila x de care ne-am ocupat pînă acuma este o variabilă de *tip discret*, ea putînd lua numai valorile 0, 1, 2, ...

Scanned with OKEN Scanner

Să considerăm de pildă populația Γ a tuturor pieselor fabricate de o mașină și fie x o anumită dimensiune a pieselor ; evident x este o variabilă de *tip continuu* definită pe această populație ; ea poate lua orice valoare dintr-un anumit interval finit. Dacă de pildă pe o populație compusă din persoane de diferite vîrste și greutăți considerăm ca variabile vîrsta sau greutatea, obținem de asemenea variabile de tip continuu. Vom determina distribuțiile unor asemenea variabile bazîndu-ne pe unele constatări deduse din experiență :

1°. Oricît de mic ar fi intervalul Δx_k în jurul valorii x_k a variabilei x, în acest interval există, în general, valori ale variabilei.

2°. În practică ne limităm la precizia pe care ne-o dau instrumentele de măsurat. Așa de pildă dacă precizia garantată este 1 mm, lungimile de 12,1 mm și 12,9 mm le considerăm egale cu 12 mm ; dacă precizia este pînă la un an, două persoane una de 24 ani și 3 zile și alta de 24 ani și 350 zile le considerăm de aceeași vîrstă. În consecință materialul statistic obținut prin selecție trebuie grupat pe anumite *intervale de variație* ale variabilei x, intervale fără puncte comune (disjuncte), de obicei, egale între ele. Modul de grupare variază, evident, în raport cu natura fenomenului sau procesului cercetat.

Să considerăm o selecție de mărime n, x_1, x_2, .., x_n și fie valorile obținute prin selecție, pe care le grupăm în r intervale egale

$$(a_1, a_2), (a_2, a_3), \ldots (a_k, a_{k+1}), \ldots (a_r, a_{r+1}) \tag{4}$$

Evident, unele din numerele a_k pot fi egale cu unele din valorile x_k $(1 \leqslant k \leqslant n)$. Intervalul (a_k, a_{k+1}) cuprinde toată mulțimea valorilor lui x pentru care avem :

$$a_k \leqslant x < a_{k+1}, \tag{4_1}$$

deci limita superioară a intervalului nu se consideră ca făcînd parte din interval. Mărimea comună a intervalelor pe care să o notăm cu d, este :

$$d = a_2 - a_1 = a_3 - a_2 = \ldots = a_{k+1} - a_k = \ldots = a_{r+1} - a_r \tag{4_2}$$

punctele $a_1, a_2, \ldots, a_{r+1}$ fiind egal distanțate pe axa absciselor, la distanța d unul de cel precedent.

Exemplu. Pe o anumită populație Γ compusă din persoane de diferite vîrste, variabila x reprezintă vîrsta. Se face o selecție de 1 000 persoane, obținînd valorile $x_1, x_2, \ldots, x_{1000}$ cuprinse

între 15 și 60 de ani. Vîrstele se grupează pe intervale de cîte 5 ani (d = 5) așa încît se formează scara de mai jos :

15—20	30—35	45—50
20—25	35—40	50—55
25—30	40—45	55—60

Avem deci $r = 9$; $a_1 = 15$, $a_2 = 20$, ... $a_9 = 55$, $a_{10} = 60$.

Grupa 45—50 de pildă, conține vîrstele 45, 46, 47, 48, 49.

Să notăm prin $n(x_k)$ numărul valorilor variabilei x care cad în intervalul (a_k, a_{k+1}). Acestea sînt *frecvențele absolute* ale acestui interval, iar raportul

$$f(x_k) = \frac{n(x_k)}{n}, \quad 1 \leqslant k \leqslant r \tag{4_3}$$

reprezintă *frecvența relativă* a aceluiași interval.

De altfel avem totdeauna relațiile cunoscute :

$$\sum_{k=1}^{r} n(x_k) = n, \qquad \sum_{k=1}^{r} f(x_k) = 1, \tag{4_4}$$

unde n este ca, de obicei, volumul selecției. Notăm prin α_k mijlocul intervalului (a_k, a_{k+1}), adică :

$$\alpha_k = \frac{a_k + a_{k+1}}{2} \tag{4_5}$$

Din această relație și din relația (4_2), $a_{k+1} - a_k = d$ deducem $a_k = \alpha_k - \frac{d}{2}$, $a_{k+1} = \alpha_k + \frac{d}{2}$, deci intervalul

$$(a_k, a_{k+1}) = \left(\alpha_k - \frac{d}{2}, \alpha_k + \frac{d}{2}\right), \tag{4_6}$$

pentru orice $k = 1,2, \ldots, r$.

Mijloacele intervalelor din exemplul dat mai sus sînt :

17,5	32,5	47,5
22,5	37,5	52,5
27,5	42,5	57,5

În acest caz, intervalul cu centrul în 37,5, de pildă, îl vom scrie sau (35,40) sau $(37{,}5 - \frac{5}{2}, 37{,}5 + \frac{5}{2})$.

Scanned with OKEN Scanner

Graficul de mai jos reprezintă scara intervalelor variabilei x sub cele două aspecte descrise mai sus :

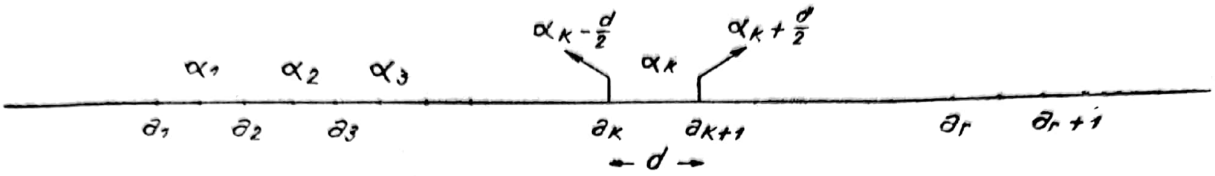

Fig. 5

Convenție. În intervalul (a_k , a_{k+1}) unele valori ale lui x sînt mai mici decît α_k, iar altele sînt mai mari decît α_k. *Toate aceste valori se consideră egale cu* α_k. Să presupunem că selecția de 1 000 persoane, amintită mai sus, se repartizează în cele 9 grupe conform următoarei situații :

Tabloul 7

Grupa vîrstelor	Numărul persoanelor
15—20	23
20—25	49
25—30	117
30—35	142
35—40	150
40—45	147
45—50	140
50—55	127
55—60	105
	1 000

De aici deducem următorul tablou de calcul :

Tabloul 8

(a_k, a_{k+1})	α_k	$n(x_k)$	$f(x_k)$	$\bar{\theta}_k$
15—20	17,5	23	0,023	2,3
20—25	22,5	49	0,049	4,9
25—30	27,5	117	0,117	11,7
30—35	32,5	142	0,142	14,2
35—40	37,5	150	0,150	15,0
40—45	42,5	147	0,147	14,7
45—50	47,5	140	0,140	14,0
50—55	52,5	127	0,127	12,7
55—60	57,5	105	0,105	10,5
		1 000	1,—	100,—

Scanned with OKEN Scanner

Datele din acest tablou se pot reprezenta grafic în mai multe feluri. Aici vom descrie numai două din ele care sînt frecvent folosite în statistica matematică.

1°. *Histograma distribuției* se obține construind dreptunghiuri pe fiecare interval avînd baza și înălțimile proporționale cu frecvențele (relative). Graficul are înfățișarea de mai jos :

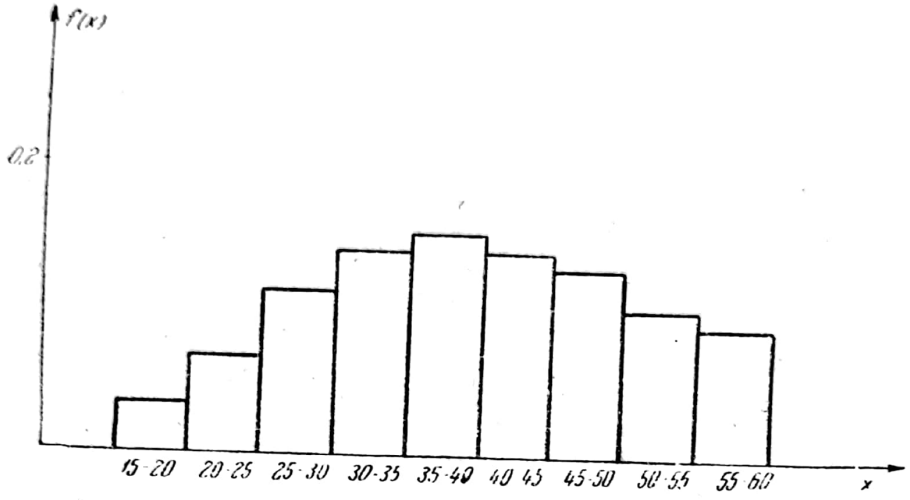

Fig. 6.

2°. *Poligonul frecvențelor* se obține ridicînd pe mijlocul fiecărui interval ordonate proporționale cu frecvențele

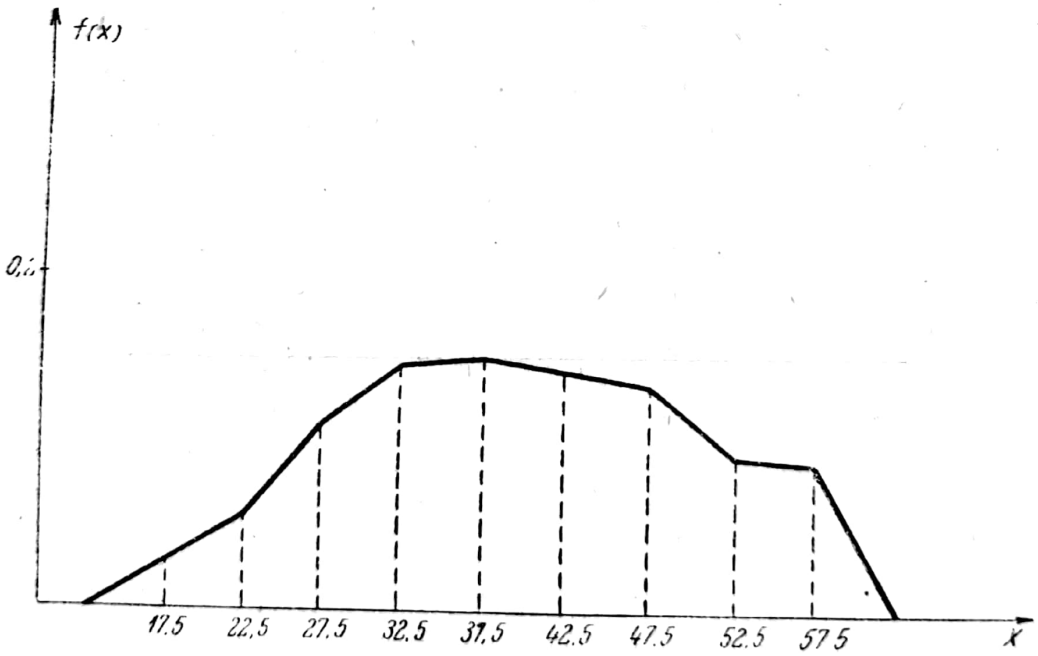

Fig. 7.

și unind apoi extremitățile acestor ordonate printr-o linie poligonală. Pentru distribuția din tabloul 8 avem poligonul de frecvență reprezentat mai sus

Scanned with OKEN Scanner

Observații. 1°. Poligonul frecvențelor se obține din histograma distribuției unind mijloacele bazelor superioare ale dreptunghiurilor ; apoi se unește mijlocul bazei superioare a primului dreptunghi cu mijlocul intervalului precedent și mijlocul bazei ultimului dreptunghi cu mijlocul intervalului situat la dreapta acestuia din urmă. Reciproc, din poligonul frecvențelor se obține histograma ducînd paralele prin vîrfuri la axa absciselor limitate pînă la intersecțiile cu ordonatele extremităților intervalelor.

2°. Ordonatele sînt proporționale cu frecvențele, deci sînt de forma $\alpha f(x_k)$; aria dreptunghiului cu baza d este α d $f(x_k)$.

Dacă alegem α așa fel ca aria acestui dreptunghi să fie $f(x_k)$, trebuie ca $\alpha = \frac{1}{d}$. Aria mărginită de histogramă este atunci

$$\sum_{k=1}^{r} f(x_k) = 1$$

În mod analog aria mărginită de poligonul frecvențelor va fi :

$$\frac{f(x_1)}{2} + \frac{f(x_1) + f(x_2)}{2} + \ldots + \frac{f(x_{r-1}) + f(x_r)}{2} + \frac{f(x_r)}{2} = \frac{2\sum_{1}^{r} f(x_k)}{2} = 1$$

Aria mărginită de histogramă pînă la ordonata punctului a_k este suma :

$$f(x_1) + f(x_2) + \ldots + f(x_{k-1}) \qquad (4_7)$$

în timp ce aria mărginită de poligonul frecvențelor pînă la aceeași ordonată este :

$$\frac{f(x_1)}{2} + \frac{f(x_1) + f(x_2)}{2} + \ldots + \frac{f(x_{k-2}) + f(x_{k-1})}{2} + \frac{\frac{1}{2}\left[f(x_{k-1}) + \frac{f(x_{k-1}) + f(x_k)}{2}\right]}{2} =$$

$$= f(x_1) + f(x_2) + \ldots + f(x_{k-1}) + \frac{f(x_{k-1}) - f(x_{k-1})}{8} \qquad (4_8)$$

Scanned with OKEN Scanner

Prin urmare, aria mărginită de histogramă este totdeauna egală cu suma frecvențelor, în timp ce aria mărginită de poligon nu este egală cu suma frecvențelor decît pentru întreaga distribuție.

Funcția de repartiție empirică a distribuției se calculează cu ajutorul formulei de mai sus

$$F_n(x) = \sum_{x_k < x} f(x_k) \tag{3}$$

De unde deducem tabloul:

$$F_n(x) = \begin{cases} 0, & \text{pentru } x < a_1 \\ f(x_1) & ,, \quad a_1 \leqslant x < a_2 \\ \sum_{k=1}^{2} f(x_k) & ,, \quad a_2 \leqslant x < a_3 \\ \cdots & \\ \sum_{k=1}^{r-1} f(x_k) & ,, \quad a_{r-1} \leqslant x < a_n \\ \sum_{k=1}^{r} f(x_k) & ,, \quad a_r \leqslant x_k \end{cases} \tag{5}$$

Pentru tabloul 8 avem valorile de mai jos:

Tabloul 9

$a_k - a_{k+1}$	$f(x_k)$	$F_n(x_k)$
15—20	0,023	0,023
20—25	0,049	0,072
25—30	0,117	0,189
30—35	0,142	0,331
35—40	0,150	0,481
40—45	0,147	0,628
45—50	0,140	0,768
50—55	0,127	0,895
55—60	0,105	1,—
	1,—	

Scanned with OKEN Scanner

Graficul funcției de repartiție este cel redat mai jos :

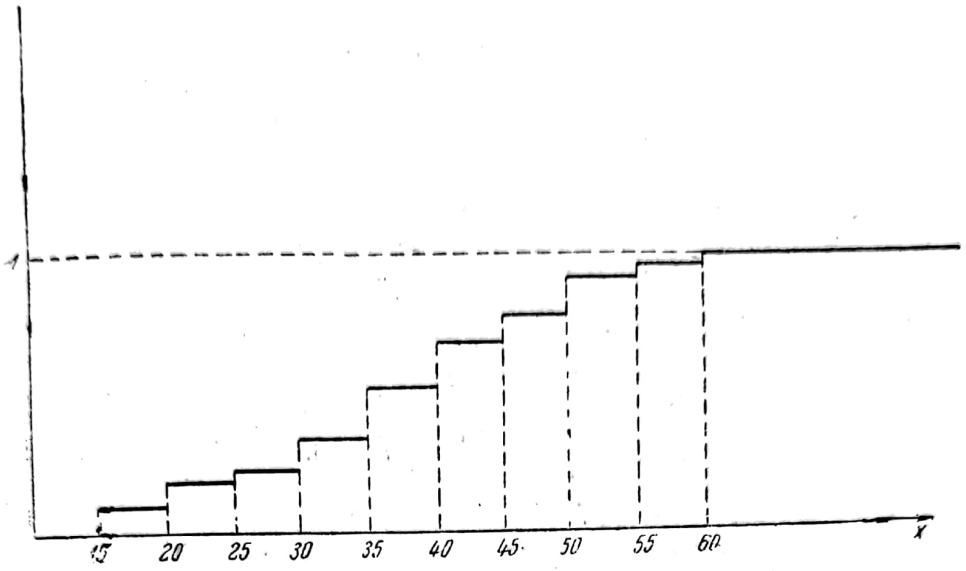

Fig. 8.

Observație. Funcția de repartiție are o interpretare geometrică simplă. În adevăr, avem prin definiție

$$F_n(a_k) = f(a_1) + f(a_2) + \ldots + f(a_{k-1}) \qquad (5_1)$$

Ținînd seama de relația (4_7), valoarea $F(a_k)$ reprezintă aria mărginită de histogramă, de axa absciselor și de ordonata punctului a_k.

§ 3. Momentele empirice ale distribuțiilor statistice

Să considerăm o selecție de mărime n obținută dintr-o populație Γ și să considerăm $x_1, x_2, \ldots, x_s$ valorile distincte între ele ale variabilei x definită în această populație. Să notăm, ca de obicei, prin $f(x_1), f(x_2) \ldots, f(x_s)$ frecvențele relative ale celor s valori. Prin definiție, *momentul de ordinul k al distribuției variabilei x* sau, mai simplu, *momentul de ordinul k al selecției* este dat de relația :

$$m'_k = \sum_{h=1}^{s} x_h^k f(x_h) \qquad (6)$$

Cînd variabila x este de tip continuu și cele n valori, $x_1, x_2, \ldots, x_n$, date de selecție se repartizează în r intervale egale cu centrele $\alpha_1, \alpha_2, \ldots, \alpha_r$, iar $f(x_1), f(x_2), \ldots, f(x_r)$ sînt frecven-

țele relative ale acestor intervale, momentul de ordinul k se definește prin relația :

$$m'_k = \sum_{h=1}^{k} \alpha_h^k f(x_h) \qquad (6_1)$$

În particular, pentru $k = 1$ obținem *valoarea medie a selecției*, care se notează de obicei prin $\bar{x}$, deci :

$$m'_1 = \bar{x} = \sum_{h=1}^{s} \alpha_h f(x_h) \qquad (6_2)$$

dacă variabila x este de tip discret și

$$m'_1 = \bar{x} = \sum_{h=1}^{r} \alpha_h f(x_h) \qquad (6_3)$$

dacă x este de tip continuu.

În mod analog se definesc și *momentele centrate ale selecției*. Vom avea pentru momentul centrat de ordinul k :

$$\mu'_k = \sum_{h=1}^{s} (x_h - \bar{x})^k f(x_h) \qquad (6_4)$$

cînd x este de tip discret și

$$\mu'_k = \sum_{h=1}^{r} (\alpha_h - \bar{x})^k f(x_h) \qquad (6_5)$$

cînd x este de tip continuu. Valorile medii $\bar{x}$ pentru (6_4) se calculează cu formula (6_2), iar pentru (6_5) cu formula (6_3)

Observații : 1°. Momentul centrat de ordinul I este nul. Să verificăm aceasta pentru (6_5). Vom putea scrie :

$$\mu'_1 = \sum_{h=1}^{r} \alpha_h f(x_h) - \bar{x} \sum_{h=1}^{r} f(x_h) = \bar{x} - \bar{x} = 0$$

după (4_4) și (6_3)

2°. Momentul centrat de ordinul II se numește *dispersia selecției* și o notă prin $\bar{\sigma}^2$, deci :

$$\mu'_2 = \bar{\sigma}^2 = \sum_{h=1}^{s} (x_h - \bar{x})^2 f(x_h) \qquad (6_6)$$

sau pentru x de tip continuu

$$\mu'_2 = \bar{\sigma}^2 = \sum_{h=1}^{r} (\alpha_h - \bar{x})^2 f(x_h) \qquad (6_7)$$

Scanned with OKEN Scanner

Avem totdeauna relația :

$$\bar{\sigma}^2 = m_2' - m_1^{12} \tag{6_8}$$

În adevăr, dezvoltînd pătratul în relația (6_6), de pildă, găsim :

$$\bar{\sigma}^2 = \sum_{h=1}^{s} x_h^2 f(x_h) - 2\bar{x} \sum_{h=1}^{s} x_h f(x_h) + \bar{x}^2 \sum_{h=1}^{s} f(x_h) =$$

$$= m_2' - 2\bar{x}^2 + \bar{x}^2 = m_2' - m_1'^2$$

Expresia :

$$\alpha = \frac{\mu_3'}{\bar{\sigma}^3} \tag{6_9}$$

ne dă indicații asupra asimetriei distribuției, iar expresia

$$\beta = \frac{\mu_4}{\bar{\sigma}^4} \tag{6_{10}}$$

ne dă indicații asupra boltirii după cum vom vedea în capitolele următoare.

3°. Între momentele centrate μ_k' și momentele m_k' avem o relație simplă. În adevăr,

$$\mu_k' = \sum_{h=1}^{s} (x_h - \bar{x})^k f(x_h) = \sum_{h=1}^{s} \left[x_h^k - C_k^1 x_h^{k-1} \bar{x} + \right.$$

$$\left. + C_k^2 x_h^{k-2} \bar{x}^2 - \ldots + (-1)^k \bar{x}^k \right] f(x_h) =$$

$$= \sum_{h=1}^{s} x_h^k f(x_h) - \bar{x} C_k^1 \sum_{h=1}^{s} x_h^{k-1} f(x_h) + C_k^2 \bar{x}^2 \sum_{h=1}^{s} x_h^{k-2} f(x_h) - \ldots +$$

$$+ (-1)^k \bar{x} \sum_{h=1}^{s} f(x_h)$$

de unde, ținînd seama de (6) obținem :

$$\mu_k' = m_k' - C_k' m', m_{k-1}' + C_k^2 m_1^{12} m_{k-2}' + \ldots + (-1)^k m_1'^k \tag{6_{11}}$$

4°. *Corecțiile lui Sheppard.* Am văzut că pentru variabilele de tip continuu, toate valorile dintr-un interval sînt înlocuite printr-un număr egal de valori și toate egale cu va-

loarea corespunzătoare mijlocului intervalului. Momentele (6_1) au fost calculate în această ipoteză care, evident, denaturează într-o oarecare măsură realitatea. Pentru aceste motive momentelor (6_1) li se aduc anumite corecții numite corecțiile lui Sheppard, pe care le vom expune într-unul din capitolele următoare.

§ 4. Medii statistice. Calculul lor simplificat

Vom începe cu mediile cele mai des întîlnite în aplicații : media aritmetică, mediana, modulul sau dominanta unei distribuții.

1°. *Media aritmetică* sau *media de selecție* este momentul de ordinul întîi al distribuției ; ea se calculează cu formulele $(6'_2)$ sau (6_3), după cum distribuția este de tip discret sau continuu. Nu ne vom ocupa aici de o serie de proprietăți ale acestei medii, ci vom enunța numai trei dintre ele care ajută fie la controlul calculului, fie să ușureze acest calcul.

Prima proprietate : Media aritmetică este o *medie internă*, adică este cuprinsă între cea mai mică și cea mai mare din valorile variabilei x. Presupnînd valorile distincte între ele $x_1, x_2, \ldots, x_s$ așezate în ordinea mărimii avem relațiile

$$x_1 < \bar{x} < x_s \tag{7}$$

În adevăr, din formula (6_2) deducem :

$$\bar{x} > \sum_{h=1}^{s} x_1 f(x_h) = x_1 \sum_{h=1}^{s} f(x_h) = x_1$$

În mod analog

$$\bar{x} < x_s \sum_{h=1}^{s} f(x_h) = x_s$$

Proprietatea a doua : Media aritmetică este *translativă*, adică satisface, oricare ar fi numărul a, relația :

$$\bar{x} + a = \sum_{h=1}^{s} (x_h + a) f(x_h) \tag{7_1}$$

Scanned with OKEN Scanner

Cu alte cuvinte, dacă mărim toate valorile cu același număr, media se mărește și ea cu acest număr. Pentru verificarea formulei (7_1) este suficient să desfacem parantezele în partea dreaptă și să ținem seama de relația $\Sigma f(x_k) = 1$.

Proprietatea a treia: Media aritmetică este o funcție omogenă și de gradul întîi de valorile selecției $x_1, x_2, \ldots, x_s$, adică:

$$\bar{x}(tx_1, tx_2, \ldots, tx_s) = t\,\bar{x}(x_1, x_2, \ldots, x_s) \qquad (7_2)$$

Proprietatea se verifică imediat înlocuind în (6_2) sau (6_3) pe x_h prin tx_h $(1 \leqslant h \leqslant s)$.

Pe proprietatea a doua se bazează calculul simplificat al mediei aritmetice. În acest scop vom alege numărul a în așa fel ca suma din partea dreaptă a relației (7_1) să se calculeze cît mai ușor posibil. Notînd prin

$$x_h - a = \xi_h \qquad (7_3)$$

valoarea ξ_h se numește *abaterea lui* x_h *de la valoarea a.*

Din (7_1) deducem:

$$\bar{x} = a + \sum_{h=1}^{s} \xi_h f(x_h) \qquad (7_4)$$

Suma $\sum_{h=1}^{s} \xi_h f(x_h)$ reprezintă media abaterilor ξ_h cu frecvențele $f(x_h)$, deci vom calcula această medie la care adăugînd a obținem media căutată $\bar{x}$. Ca aplicație să calculăm media artimetică a distribuției din tabloul 1. Vom lua în acest caz pentru a valoarea centrală 3. Avem tabloul de calcul:

Tabloul 10

x_h	$\xi_h = x_h - 3$	$f(x_h)$	$\xi_h \cdot f(x_h)$
0	−3	0,165	−0,495
1	−2	0,322	−0,644
2	−1	0,244	−0,244
3	0	0,135	−1,383
4	+1	0,073	0,073
5	+2	0,040	0,080
6	+3	0,021	0,063
		1,—	+0,216

De aici deducem media abaterilor :

$$\sum_{h=0}^{6} \xi_h f(x_h) = 0{,}216 - 1{,}383 = -1{,}167$$

de unde media : $\bar{x} = 3 + (-1{,}167) = 1{,}833 \approx 2$

Observație

La aceeași valoare ajungem aplicînd direct formula (6_2). Vom avea :

Tabloul 11

x_h	$f(x_h)$	$x_h f(x_h)$
0	0,165	—
1	0,322	0,322
2	0,244	0,488
3	0,135	0,405
4	0,073	0,292
5	0,040	0,200
6	0,021	0,126
	1,—	1,833

Să calculăm acum media aritmetică a distribuției din tabloul 7

Astfel, aplicînd formula (6_3) vom obține :

Tabloul 12

α_h	$f(x_h)$	$\alpha_h f(x_h)$
17,5	0,023	0,4025
22,5	0,049	1,1025
27,5	0,117	3,2175
32,5	0,142	4,6150
37,5	0,150	5,6250
42,5	0,147	6,2475
47,5	0,140	6,6500
52,5	0,127	6,6675
57,5	0,105	6,0375
	1,—	$40{,}5650 = \bar{x}$

Scanned with OKEN Scanner

Folosind calculul simplificat, vom lua valoarea a = 37,5 și obținem :

Tabloul 13

α_h	$\xi_h = \alpha_h - 37{,}5$	$\xi(x_h)$	$\xi_h f(x_h)$
17,5	−20	0,023	−0,460
22,5	−15	0,049	−0,735
27,5	−10	0,117	−1,170
32,5	−5	0,142	−0,710
37,5	0	0,150	−3,075
42,5	+ 5	0,147	0,735
47,5	+10	0,140	1,400
52,5	+15	0,127	1,905
57,5	+20	0,105	2,100
		1,—	6,140

De unde rezultă :

$$\sum \xi_h f(x_h) = 6{,}140 - 3{,}075 = 3{,}065$$

$$\bar{x} = 37{,}5 + 3{,}065 = 40{,}565 \text{ ani}$$

Acest calcul mai poate fi încă simplificat folosind proprietatea III ; în acest scop, vom împărți coloana ξ_h prin lungimea d = 5 a intervalului, avînd grijă ca după efectuarea calculelor să înmulțim rezultatul cu acest număr.

Tabloul 14

α_h	$\alpha_h = d_h - 37{,}5$	$\beta_h = \xi_h/5$	$f(x_h)$	$\beta_h f(x_h)$
17,5	−20	−4	0,023	−0,092
22,5	−15	−3	0,049	−0,147
27,5	−10	−2	0,117	−0,234
32,5	−5	−1	0,142	−0,142
37,5	0	0	0,150	−0,615
42,5	+ 5	+1	0,147	0,147
47,5	+10	+2	0,140	0,280
52,5	+15	+3	0,127	0,381
57,5	+20	+4	0,105	0,420
			1,—	1,228

De aici deducem media : $\bar{x} = 37{,}5 + (1{,}228 - 0{,}615) \cdot 5 =$ 40,565 ani

Scanned with OKEN Scanner

2°. *Mediana — Me —* a distribuției este valoarea lui x pentru care avem relația :

$$F_n(Me) = \frac{1}{2} \tag{8}$$

unde $F_n(x)$ este funcția de repartiție.
De aici rezultă că dacă x este de tip discret mediana va fi valoarea x_k pentru care :

$$F_n(x_k) \geqslant \frac{1}{2} \tag{8_1}$$

Pentru tabloul 1 avem mediana Me = 2, deoarece $F_n(1) = 0{,}487 < \frac{1}{2}$, $F_n(2) = 0{,}731 > \frac{1}{2}$ (Tabloul 4). Mediana acestei distribuții se poate obține și geometric intersectînd graficul funcției $F_n(x)$ cu paralela la axa absciselor care trece prin punctul de ordonată 0,5 de pe axa ordonatelor. Această paralelă, se vede din fig. 2 că întîlnește graficul pe ordonata punctului x = 2.

Calculul medianei pentru distribuții continue este ceva mai complicat. Vom stabili o formulă aproximativă folosind interpretarea geometrică dată funcției de repartiție (relația 5_1)

Fie (a_k, a_{k+1}) intervalul pentru care avem :

$$F_n(a_k) < \frac{1}{2},\ F_n(a_{k+1}) > \frac{1}{2} \tag{8_2}$$

Mediana va fi egală cu $a_k + x'$, unde x' este distanța de la punctul a_k pînă la ordonata punctului Me care împarte aria mărginită de histogramă în două părți egale.

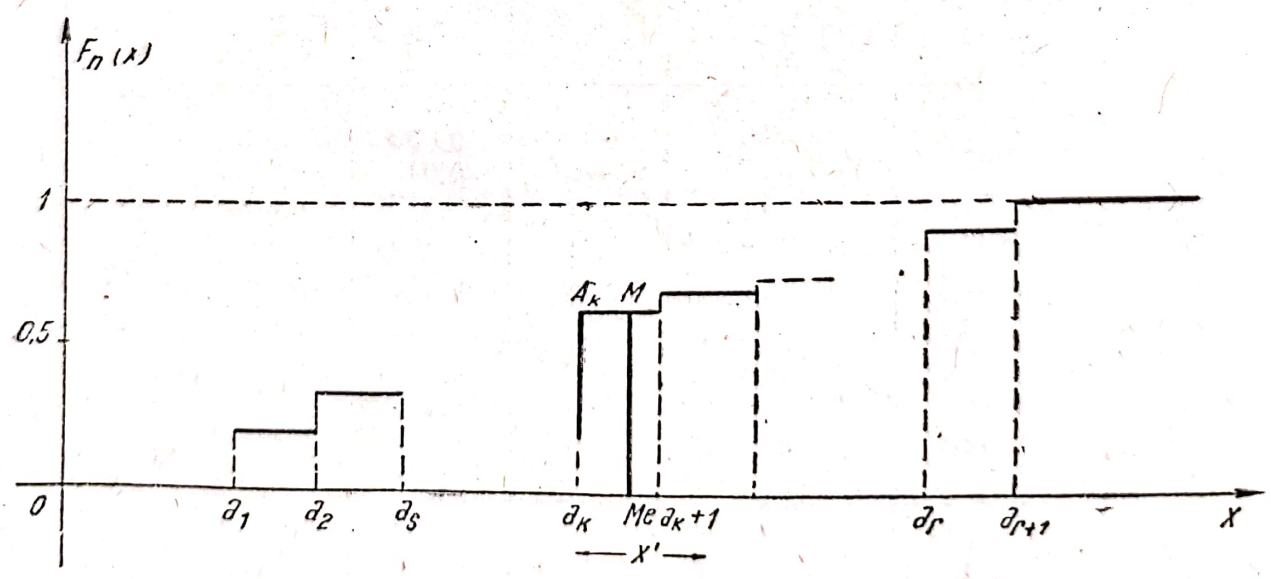

Fig. 9.

Aria mărginită de histogramă pînă la ordonata punctului M_e este egală cu $\frac{1}{2}$ prin definiție, iar aria pînă la ordonata lui a_k este egală cu $F_n(a_k)$, după relația (5_1). Diferența lor este, aria dreptunghiului $a_k A_k M M_e$ cu baza x' și înălțimea $\frac{f(x_k)}{d}$ deci :

$$x' \frac{f(a_k)}{d} = \frac{1}{2} - F_n(a_k)$$

De unde deducem :

$$x' = \frac{\frac{1}{2} - F_n(a_k)}{f(x_k)/d} = \frac{1 - 2F_n(a_k)}{2f(x_k)} d$$

Mediana se va calcula deci cu ajutorul formulei

$$Me = a_k + \frac{1 - 2F_n(a_k)}{2f(x_k)} d \qquad (8_3).$$

Exemplu : Să calculăm mediana distribuției din tabloul 7. Folosind datele tabelului 9, vom obține :

$$F_n(40) = 0{,}481 < \frac{1}{2}, \quad F_n(45) = 0{,}628 > \frac{1}{2}$$

Deci, mediana se găsește în intervalul $(40 - 45)$:

$$a_k = 40, \qquad a_{k+1} = 45$$

Deoarece $d = 5$, $f(x_k) = f(40) = 0{,}147$ obținem :

$$Me = 40 + \frac{1 - 2 \times 0{,}481}{2 \times 0{,}147} \cdot 5 = 40{,}65$$

3°. *Modul* M_o al distribuției este valoarea lui x pentru care *frecvența* $f(x)$ *este maximă*. De aici rezultă că pentru distribuțiile discrete, în tabel vom citi frecvența maximă și valoarea corespunzătoare a lui x este modulul distribuției. Așa de pildă pentru distribuția din tabloul 1 avem $M_o = 1$, deoarece frecvența cea mai mare este 0,322 căreia îi corespunde $x = 1$. Dacă distribuția este continuă, se poate lua $M_o = \alpha_k$, unde α_k este mijlocul intervalului (a_k, a_{k+1}) pentru care frecvența $f(x_k)$ are valoarea cea mai mare. Pentru distribuția din tabelul 7 putem lua $M_o = 37{,}5$ deoarece intervalul $35 - 40$ are frecvența cea mai mare.

Scanned with OKEN Scanner

Se obține însă o valoare mai exactă a modului unei distribuții continue folosind și frecvențele intervalelor alăturate intervalului modal (cel cu frecvența maximă). În acest scop, linia poligonală $A_{k-1}A_k\,A_{k+1}$ se înlocuiește printr-o linie curbă și anume printr-o parabolă de ordinul al doilea.

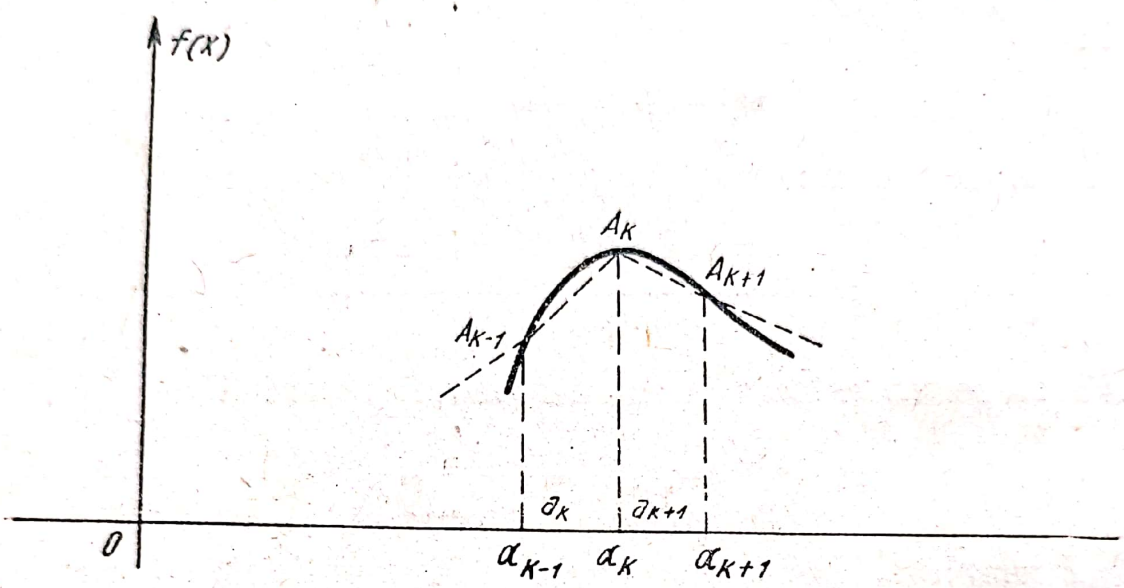

Fig. 10.

Luînd ca origine punctul α_k mijlocul intervalului (a_k, a_{k+1}), pentru care frecvența $f(x_k)$ este maximă, ecuația acestei parabole este de forma : $f(x) = ax^2 + bx + c$.

Presupunînd că pentru $\alpha_{k-1} = \alpha_k - d$ avem $x = -d$ (cînd luăm ca origine pe α_k), ordonata acestui punct este $f(x_{k-1})$, pentru α_k $(x = 0)$ avem ordonata $f(x_k)$, iar pentru $\alpha_{k+1} = a_k + d$, cînd $x = d$, avem ordonata $f(x_{k+1})$, obținem sistemul în a, b, c :

$$ad^2 - bd + c = f(x_{k-1})$$
$$c = f(x_k)$$
$$ad^2 + bd + c = f(x_{k+1})$$

Rezolvînd acest sistem deducem valorile :

$$a = \frac{f(x_{k-1}) - 2f(x_k) + f(x_{k+1})}{2d^2}$$

$$b = \frac{f(x_{k+1}) - f(x_{k-1})}{2d}$$

$$c = f(x_k)$$

Maximul funcției f (x) se obține din relația : $f'(x) = 2ax + b = 0$
De unde deducem valoarea :

$$x = -\frac{b}{2a} = -\frac{f(x_{k+1}) - f(x_{k-1})}{f(x_{k-1}) - 2f(x_k) + f(x_{k+1})} \cdot \frac{d}{2}$$

Așadar pentru distribuțiile continue, modul se calculează cu formula :

$$M_0 = \alpha_k - \frac{f(x_{k+1}) - f(x_{k-1})}{f(x_{k-1}) - 2f(x_k) + f(x_{k+1})} \cdot \frac{d}{2} \qquad (9)$$

unde α_k este mijlocul intervalului modal.

Pentru distribuția din tabloul 7 avem :

$$\alpha_k = 37{,}5 \qquad d = 5$$

$$f(x_{k-1}) = 0{,}142, \qquad f(x_k) = 0{,}150, \qquad f(x_{k+1}) = 0{,}147$$

de unde :

$$M_0 = 37{,}5 - \frac{0{,}147 - 0{,}142}{0{,}147 - 2 \times 0{,}150 + 0{,}142} \cdot \frac{5}{2} = 38{,}6$$

Medii de ordin superior. Să luăm, de pildă, $\varphi(x)$ o funcție definită pe mulțimea Γ și $x_1, x_2, \ldots, x_n$ valorile obținute printr-o selecție de mărime n efectuată din această mulțime.

Dacă numai s din ele sînt distincte, fie $x_1, x_2, \ldots, x_s$ iar $f(x_1)$, $f(x_2)$, ..., $f(x_s)$ frecvențele relative corespunzătoare. O clasă importantă de medii se obține din ecuația

$$\varphi(m) = \sum_{h=1}^{s} \varphi(x_h) f(x_h) \qquad (10)$$

unde cu m am însemnat media

$$m = m(x_1, x_2, \ldots, x_s) \qquad (10_1)$$

Dacă $\varphi(x) = x$, obținem următoarea medie aritmetică

$$m = \sum_{h=1}^{s} x_h f(x_h) \qquad (10_2)$$

și în particular, cînd valorile selecției sînt distincte $f(x_h) = \frac{1}{n}$,

$$m = \frac{1}{n} \sum_{h=1}^{s} x_h \qquad (10'_2)$$

Scanned with OKEN Scanner

Dacă $\varphi(x) = \frac{1}{x}$ obținem media armonică:

$$\frac{1}{m} = \sum_{h=1}^{s} \frac{1}{x_h} f(x_h)$$

sau încă

$$m = \frac{1}{\sum_{h=1}^{s} \frac{f(x_h)}{x_h}} \qquad (10_3)$$

Cînd valorile sînt distincte obținem:

$$m = \frac{n}{\sum_{h=1}^{s} \frac{1}{x_h}} \qquad (10'_3)$$

Pentru $\varphi(x) = \ln x$ se găsește media:

$$\ln m = \sum_{h=1}^{s} f(x_h) \ln x_h = \sum_{h=1}^{s} \ln x_h^{f(x_h)} = \ln \prod_{h=1}^{s} x_h^{f(x_h)}$$

de unde se obține media geometrică

$$m = x_1^{f(x_1)} x_2^{f(x_2)} \dots x_s^{f(x_s)} \qquad (10)_4$$

iar dacă valorile sînt distincte

$$m = \sqrt[n]{x_1 x_2 \dots x_n} \qquad (10'_4)$$

În general, dacă $\varphi(x) = x^k$ obținem media de ordinul k:

$$m^k = \sum_{h=1}^{s} x_h^k f(x_h)$$

de unde:

$$m = \sqrt[k]{\sum_{h=1}^{s} x_h^k f(x_h)} \qquad (10_5)$$

Pentru valorile distincte ale selecției avem:

$$m = \sqrt[k]{\frac{1}{n} \sum_{h=1}^{n} x_h^k} \qquad (10'_5)$$

Scanned with OKEN Scanner

Observație : În unele chestiuni de ordin teoretic este mai simplu să folosim pentru medii expresiile simplificate cînd toate valorile selecției sînt diferite între ele. Acest procedeu nu restrînge întru nimic generalitatea problemei deoarece de la o medie simplă putem trece la o medie ponderată grupînd într-un singur termen valorile egale şi reciproc, de la o medie ponderată putem trece la una simplă scriind un termen de atîtea ori de cîte unități conține frecvența sa absolută.

Relația de ordine între diferite medii. Dacă vom nota prin $m_{(k)}$ media de ordinul k, dată de relația (10_5), prin particulari- zarea indicelui k obținem toate mediile descrise mai sus ca pe niște cazuri particulare.

În adevăr, pentru $k=1$ obținem media aritmetică, iar pentru $k=2$ avem media pătratică

$$m_{(2)} = \sqrt{\frac{1}{n}\sum_{h=1}^{n} x_h} \qquad (10_6)$$

Pentru $k = -1$ se găsește media armonică. În adevăr, putem scrie media dată de relația 10_5) sub forma :

$$m_{(k)} = \sqrt[k]{\frac{\sum_{h=1}^{n} x_k^h}{n}} = \left(\frac{\sum_{h=1}^{n} x_h^k}{n}\right)^{\frac{1}{k}} \qquad (10_5'')$$

de unde în cazul cînd $k = -1$ obținem :

$$m_{(-1)} = \left(\frac{\sum_{h=1}^{n} \frac{1}{x_h}}{n}\right)^{-1} \frac{n}{\sum_{h=1}^{n} \frac{1}{x_h}}$$

Cînd $k \to 0$ obținem media geometrică. Pentru aceasta vom face relația în $(10_5'')$ pe $k = 0$ şi obținem 1^{∞}. Aplicînd logaritmii se găsește :

$$\ln m_{(k)} = \frac{1}{k} \ln \frac{\sum_{h=1}^{n} x_h^k}{n}$$

Scanned with OKEN Scanner

De unde, cu ajutorul regulii Hôspital, obținem :

$$\lim_{k \to 0} \ln m_{(k)} = \lim_{k \to 0} \frac{\frac{1}{n} \sum_{h=1}^{n} x_k^h \ln x_k}{\frac{1}{n} \sum_{h=1}^{n} x_n^k} = \frac{1}{n} \sum_{h=1}^{n} \ln x_h =$$

$$= \ln (x_1 \cdot x_2 \cdot x_3 \ldots x_n)^{\frac{1}{n}}$$

$$m_{(0)} = \sqrt[n]{x_1 \cdot x_2 \ldots x_n}$$

adică tocmai media geometrică.

Teoremă. Între mediile m_k și m_r avem relația :

$$m_{(k)} \leqslant m_{(r)} \qquad (10_7)$$

pentru orice $k \leqslant r$ și $x_h > 0$ $(h = 1, 2, \ldots, n)$.

În acest scop vom demonstra mai întîi inegalitatea

$$f\left(\sum_{h=1}^{n} \lambda_h x_h \right) \leqslant \sum_{h=1}^{n} \lambda_h f(x_h) \qquad (10_8)$$

în care $x_1, x_2, \ldots, x_n$ sînt cuprinse într-un interval (a, b), numerele $\lambda_1, \lambda_2, \ldots, \lambda_n$ sînt cuprinse între 0 și 1 așa fel ca

$\sum_{h=1}^{n} \lambda_h = 1$, iar funcția $f(x)$ este convexă în (a, b).

Funcția f (x) este convexă cînd pentru orice pereche de numere $x', x'' \in (a, b)$[1] subsistă relația :

$$f[\lambda x' + (1-\lambda)] x'' \leqslant \lambda f(x') + (1-\lambda) f(x'') \qquad (10_9)$$

Vom demonstra relația (10_9) prin recurență.

Dacă $n = 2$ avem relația (10_8). Admitem forma (10_8) pentru $(n-1)$ și s-o demonstrăm pentru n valori.

Să considerăm numărul $k = \sum_{h=1}^{n-1} \lambda_h > 0$ și punctul

$$x = \frac{\sum_{h=1}^{n-1} \lambda_h x_h}{\sum_{h=1}^{n-1} \lambda_h}$$

[1] Arătăm în scris că numărul x aparține intervalului (a, b) astfel : $x \varepsilon (a,b)$ Semnul ε se numește semnul „apartenenței".

Înlocuind x_h ($h = 1, 2, \ldots, n$) prin a și b deducem : $a \leqslant x \leqslant b$ deci $x \in (a, b)$. În baza relației (10_8) avem pentru $(n-1)$

$$f\left(\sum_{h=1}^{n-1} \frac{\lambda_h}{h}\, x_h\right) \leqslant \sum_{h=1}^{n} \frac{\lambda_h}{k}\, f(x_h)$$

sau

$$k\, f(x) \leqslant \sum_{h=1}^{n-1} \lambda_h\, f(x_h)$$

În baza relației (10_9) avem însă

$$f\left(\sum_{h=1}^{n} \lambda_n\, x_h\right) = f\,[kx + (1-k)\, x_n] \leqslant k\, f(x) + (1-k)\, f(x_n) \leqslant$$

$$\leqslant \sum_{h=1}^{n} \lambda_h\, f(x_n)$$

adică ceea ce trebuia să demonstrăm.

Dacă $\lambda_1 = \lambda_2 = \ldots = \lambda_n = \frac{s}{n}$ din relația (10_8) deducem :

$$f\left(\frac{1}{n} \sum_{h=1}^{n} x_h\right) \leqslant \frac{1}{n} \sum_{h=1}^{n} f(x_h)$$

Dacă $f(x) = x^k$ funcția $f(x)$ este convexă pentru $k > 1$ și $x > 0$, deci

$$\left(\frac{x_1 + x_2 + \ldots + x_n}{n}\right)^k \leqslant \frac{x_1^k + x_2^k + \ldots + x_n^k}{n}$$

Inegalitatea se păstrează dacă înlocuim pe x_h prin x_h^r deci,

$$\left(\frac{1}{n} \sum x_h^r\right)^k \leqslant \frac{1}{n} \sum x_h^{kr}$$

Ridicînd apoi la puterea $\frac{1}{rk}$ găsim :

$$\left(\frac{1}{n} \sum x_h^r\right)^{\frac{1}{r}} \leqslant \left(\frac{1}{n} \sum x_h^{rk}\right)^{\frac{1}{rh}}$$

Așadar pentru $k > 1$ și $r > 0$ avem $m_{(r)} < m_{(rk)}$. Dacă $r < 0$, înlocuind x_h prin x_h^{-r} și ridicînd la puterea $-\frac{1}{rk}$ găsim $m_{(-r)} \geqslant m_{(-r)}$, deci în toate cazurile subsistă relația (10_7).

În particular putem afirma că media armonică este cel mult egală cu cea geometrică, iar aceasta la rîndul ei este cel mult egală cu media aritmetică.

Observații. Mediile de diferite ordine descrise mai înainte nu se aplică uniform în toate cazurile concrete. Așa de pildă într-o problemă este indicat să se folosească media aritmetică, în alta media armonică etc.

În acest scop, vom da următoarea definiție a valorii medii, cunoscută sub numele de definiția Chisini-Boiarski :

1°. Se numește *proprietate determinantă* a populației Γ, acea proprietate legată de populația Γ care trebuie să rămînă neschimbată, oricare ar fi variațiile posibile ale valorilor x_k $(1 \leqslant k \leqslant s)$ ale selecției.

2°. Se numește *valoare medie* a variabilei x definită pe Γ, după proprietatea determinantă, acea valoare $\bar{x}$ care prin substituția

$$x_k = \bar{x} \quad (x = 1, 2, \ldots, s)$$

nu modifică proprietatea determinantă.

În baza definiției 1°, deducem că proprietatea determinantă este o funcție de valorile $x_1, x_2, \ldots x_s$, fie $F(x_1, x_2, \ldots x_s)$; după proprietatea 2°, media $\bar{x}$ se determină din ecuația

$$F(x_1, x_2, \ldots, x_s) = F(\bar{x}, \bar{x}, \ldots \bar{x}) \tag{11}$$

Exemplul 1. Aliajele $A_1, A_2, \ldots, A_s$ au respectiv titlurile $t_1, t_2, \ldots, t_s$ și greutățile $g_1, g_2, \ldots g_s$. Să se calculeze titlul mediu al aliajului obținut topind împreună cele n_s aliaje

Luăm ca proprietate determinantă cantitatea de metal prețios. Atunci t_k reprezintă cantitatea de metal prețios dintr-un gram din aliajele A_k $(1 \leqslant k \leqslant s)$ și, deoarece în acest aliaj sînt g_k grame, vom avea $g_k t_k$ grame metal prețios.

Putem deci scrie : $F(t_1, t_2, \ldots, t_s) = g_1 t_1 + g_2 t_2 + \ldots + g_s t_s$

Media căutată $\bar{t}$ verifică ecuația $F(t_1, t_2, \ldots, t_s) = F(t, t, \ldots, t)$ adică :

$$g_1 \bar{t} + g_2 \bar{t} + \ldots g_s \bar{t} = \sum_{h=1}^{s} g_k t_k$$

De unde deducem media aritmetică

$$t - \frac{\sum_{k=1}^{s} g_k t_k}{\sum_{k=1}^{s} g_k}$$

Scanned with OKEN Scanner

Exemplul 2. În unitatea de timp, o mașină consumă o cantitate de benzină proporțională cu pătratul vitezei sale. Pe intervalele de timp $x_1, x_2, \ldots, x_n$ mașina merge cu vitezele $v_1, v_2, \ldots, v_n$. Să se determine viteza medie ca în timpul $\sum x_k$ să se consume aceeași cantitate de benzină.

Proprietatea determinantă este cantitatea de benzină consumată. Variabila ia valorile $v_1, v_2, \ldots, v_n$ deci:

$$F(v_1, v_2, \ldots, v_n) = \sum_{k=1}^{n} q\, v_k^2 x_k$$

unde q este un factor de proporționalitate. Vom avea deci ecuația

$$q \sum_{k=1}^{n} v_k^2 x_k = q \sum_{k=1}^{n} \bar{v}^2 x_k$$

de unde rezultă media:

$$\bar{v} = \sqrt{\frac{\sum v_h^2 x_k}{x_k}}$$

adică o medie de ordinul II.

Exemplul 3. În lunile aprilie și mai prețul unei mărfi a fost p_1, iar în luna iunie s-a redus la p_2. Știind că în primele două luni s-a vîndut marfă în valoare de S_1 lei, iar în iunie s-a vîndut de S_2 lei, să se determine prețul mediu al acestei mărfi pe trimestrul aprilie-iunie.

Proprietatea determinantă este cantitatea de marfă vîndută în acest trimestru, iar variabila este prețul, astfel: $F(p_1, p_2) = Q_1 + Q_2$ unde Q_1 este cantitatea vîndută în primele două luni, iar Q_2 este cantitatea vîndută în luna treia. Avem însă ecuația:

$$Q_1 = \frac{S_1}{p_1}, \quad Q_2 = \frac{S_2}{p_2}$$

de unde:

$$F(p_1, p_2) = \frac{S_1}{p_1} + \frac{S_2}{p_2}$$

Prețul mediu p rezultă din ecuația

$$\frac{S_1}{p_1} + \frac{S_2}{p_2} = \frac{S_1}{\bar{p}} + \frac{S_2}{\bar{p}}$$

Scanned with OKEN Scanner

Din această ecuație deducem :

$$\bar{p} = \frac{S_1 + S_2}{\frac{S_1}{p_1} + \frac{S_2}{p_2}}$$

deci o medie armonică.

Probleme

1. Se dă tabloul :

x	5	6	7	8	9	10	11	12	13	14	15
n(x)	1	5	9	22	38	62	61	29	24	4	1

Să se calculeze frecvențele relative și procentele corespunzătoare.

2. Pentru distribuția de mai jos :

x	n (x)
25—30	1 391
30—35	2 942
35—40	3 105
40—45	2 754
45—50	2 390
50—55	1 729
55—60	1 384

să se calculeze frecvențele relative, procentele și promilele.

3. Să se reprezinte grafic datele din problema 1 și funcția de repartiție a distribuției.

4. Aceeași problemă pentru distribuția (2).

5. Să se calculeze media aritmetică, geometrică și armonică pentru distribuția (1) și să se compare rezultatele.

6. Aceeași problemă pentru distribuția (2).

7. Să se calculeze mediile distribuțiilor (1) și (2) folosind valoarea 12 și respectiv intervalul 35—40.

8. Să se determine grafic și prin calcul mediana și modul distribuțiilor (1) și (2).

9. Fie $\Gamma_1, \Gamma_2, \ldots, \Gamma_s$, pentru s selecții din populația Γ cu volumele $n_1, n_2, \ldots, n_s$ și mediile $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_s$. Să se arate că media ansamblului de s selecții este dată de relația

$$\bar{x} = \frac{\sum n_k \bar{x}_k}{\sum n_k}$$

10. Să se calculeze dispersia distribuției (1').

Scanned with OKEN Scanner

11. Aceeași problemă pentru distribuția (2).

12. Să se demonstreze că media contra armonică

$$\sum x_k^2 : \sum x_k$$

este o medie internă.

13. Raportul dintre lungimea unui fir (în metri) și greutatea lui (în grame) se numește în industria textilă numărul tortului.

Pentru un fir tors pe porțiunile de lungimi $l_1, l_2, \ldots, l_r$ s-au calculat numerele $n_1, n_2, \ldots n_r$.

Să se calculeze numărul mediu pentru firul întreg.

14. Să se demonstreze identitatea:

$$\sum_{k=1}^{n} (x_k - \bar{x})^2 = \sum_{k=1}^{n} x_k^2 - \frac{1}{n} \left(\sum_{k=1}^{n} x_k \right)^2$$

unde $\bar{x} = \frac{1}{n} (x_1 + x_2 + \ldots + x_n)$.

15. Pentru distribuțiile (1) și (2) să se calculeze expresiile $\frac{\mu_3'}{\bar{\sigma}^3}$ și $\frac{\mu_4'}{\bar{\sigma}^4}$ unde $\bar{\sigma} = \sqrt{\mu_2}$

Capitolul II

NOȚIUNI DE BAZĂ ÎN TEORIA PROBABILITĂȚILOR

Întrucît statistica matematică studiază procesele de masă cu ajutorul calculului probabilităților, vom începe studiul de față cu lămurirea noțiunilor de bază din teoria probabilităților.

A. Probe. Realizarea în mod practic a unui complex de condiții se numește *probă*. Acest complex de condiții pe care îl vom preciza de fiecare dată îl însemnăm în general prin Γ.

Exemple: 1°. Se așază în paralel un ac magnetic cu un conductor electric prin care trece curentul. Complexul Γ constă din această operație, deci ea constituie o probă în sensul definit mai sus.

2°. Fierul topit se amestecă cu carbon în proporție de 1,7%. Această operație este o probă.

3°. Dintr-un lot de două becuri de aceeași formă și culoare, unul avînd 60 de lumini și celălalt 75 de lumini, se extrage la întîmplare unul din ele. Complexul Γ constă din această extracție : ea constituie deci o probă.

4°. Se aruncă un zar, perfect construit din punct de vedere geometric și omogen din punct de vedere fizic. Complexul Γ constă din această operație.

5°. Într-un lot de 17 piese, 15 corespund condițiilor standard și două nu corespund. Extracția la întîmplare a unei piese din acest lot constituie o probă.

6°. Alegerea la întîmplare a unui punct situat în intervalul (0,1) este o probă.

B. Evenimente. Fiecare din probele de mai sus conduce la un anumit rezultat, numit *eveniment*. În proba 1°, evenimentul constă din devierea acului magnetic; în proba 2° se obține fonta. În proba 3° evenimentul constă fie în obținerea becului de 60 de lumini, fie a celui de 75 de lumini. În proba 4° eveni-

Scanned with OKEN Scanner

mentul constă în apariția sau a feței 1, ..., sau a feței 6. În proba 5°, putem obține fiecare din cele 15 piese standard sau fiecare din cele 2 piese rebut. În sfîrșit, în proba 6° evenimentul poate consta fie din alegerea unui punct rațional (adică un punct avînd abscisa număr rațional), fie a unui punct irațional. Fiecare din acestea sînt în număr infinit.

Observații: I. Din aceste exemple deducem o primă clasificare a evenimentelor. Așa de pildă în 1° și 2° o dată cu realizarea complexului Γ obținem devierea acului magnetic; respectiv fonta. Asemenea evenimente se numesc *evenimente sigure* și teoria probabilităților nu se ocupă cu cercetarea lor. În probele 3° — 6° însă, rezultatul fiecăreia depinde de întîmplare. Unor astfel de evenimente li se spune *evenimente aleatorii* și ele formează obiectele de bază ale teoriei probabilităților.

II. Fiecare din probele 3° — 6° au fost executate asupra unei mulțimi de elemente identice ca formă, deosebindu-se numai prin anumite caracteristici pe care le constatăm după efectuarea probei. Aceste mulțimi se numesc *populații* sau *colectivități*.

Numărul elementelor care compun colectivitatea formează *volumul colectivității*.

În exemplul 3° volumul este egal cu 2, în 4° cu 6, în 5° cu 17, iar în exemplul 6°, volumul colectivității este format din infinitatea continuă de numere cuprinsă în intervalul deschis (0,1). În condițiile arătate pînă acuma, o probă poate deci conduce la oricare din elementele unei mulțimi finite sau infinite.

C. Cîmp de evenimente. Noțiunile care urmează sînt cele cunoscute din teoria mulțimilor, ale căror elemente sînt acum evenimente aleatorii. Pentru o mai bună înțelegere a lor vom folosi adesea unele reprezentări geometrice.

Să considerăm astfel pătratul din fig. 11 și să notăm prin E evenimentul care constă în aceea că un punct interior aparține acestui pătrat.

Vom numi acest eveniment, *evenimentul sigur*; un punct exterior pătratului de pildă, va fi evenimentul imposibil, deoarece el nu poate aparține pătratului.

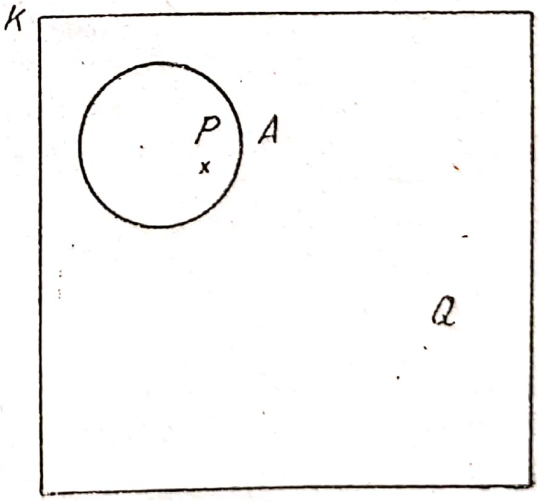

Fig. 11.

Notăm evenimentul imposibil prin Φ. Faptul că un punct oarecare al pătratului este interior cercului A constituie *evenimentul* A.

Deoarece realizarea evenimentului A (adică punctul P se găseşte în cercul A), atrage realizarea lui E (adică punctul P aparţine şi pătratului [1]), vom spune că evenimentul A implică evenimentul E şi scriem :

$$A \supset E \text{ sau } E \subset A \qquad (1)$$

Relaţia întîia se citeşte „A implică E", iar a doua „E implicat de A". Cele două relaţii care exprimă acelaşi lucru se zic *echivalente*.

Punctul Q al pătratului nu aparţine cercului A, adică $Q \in K$, dar $Q \in A$. Acest fapt constituie un nou eveniment care constă în nerealizarea evenimentului A. El reprezintă mulţimea punctelor pătratului exterioare cercului A. Aşadar, dacă facem o probă alegînd un punct al lui K la întîmplare, acest punct sau se găseşte în A sau este exterior lui A. Notînd acest din urmă eveniment prin A^*, vom spune că evenimentele A şi A^* sînt *evenimente contrare*.

Într-o probă, dacă se realizează A nu se mai realizează şi A^* şi reciproc. Evenimentul contrar lui A îl vom nota simbolic şi E — A, adică :

$$A^* = E - A \qquad (1_1)$$

înţelegînd prin aceasta mulţimea tuturor punctelor pătratului cu excepţia celor din cercul A. Este evident că şi $A^* \subset E$.

Dacă B este alt eveniment şi dacă avem : $A \subset B$ şi $B \subset A$ rezultă $A = B$. Primele două relaţii sînt echivalente cu ultima. În interpretarea geometrică pe care am dat-o evenimentelor, prima relaţie exprimă faptul că cercul A este interior lui B ; a doua exprimă din contră faptul că B este în acelaşi timp interior lui A, deci cercurile A şi B coincid $A = B$.

Reuniunea evenimentelor A, B însemnează realizarea sau a evenimentului A sau a evenimentului B, indiferent care din ele. Obţinem astfel un nou eveniment S care se notează astfel :

$$S = A \cup B \qquad (2)$$

Se citeşte *S egal A reunit cu B*. Dacă evenimentele A şi B sînt reprezentate prin cercurile A şi B din fig. 12, reuniunea lor este

---

[1] Arătăm în scris că punctul P aparţine pătratului K astfel : $P \varepsilon K$ şi citim „P aparţine lui K". Dacă P este exterior scriem $P \bar{\varepsilon} K$ şi citim P nu aparţine lui K

reprezentată prin interiorul haşurat al celor două cercuri. Cu alte cuvinte, faptul că un punct oarecare al pătratului K se găseşte în una din astcee regiuni constituie evenimentul $A \cup B$. Dacă cercurile se intersectează ca în fig. 13, reuniunea lor este reprezentată prin partea haşurată.

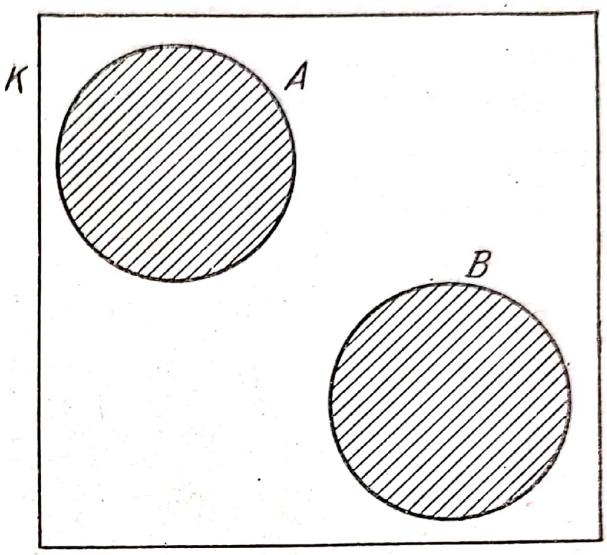

Fig. 12

Deoarece realizarea fie a lui A fie a lui B atrage realizarea evenimentului $A \cup B$ avem :

$$A \subset A \cup B, \qquad B \subset A \cup B \quad (2_1)$$

Cazuri particulare.

1°. Dacă $A \subset B$, atunci $A \cup B = B$. Geometric, aceasta înseamnă că cercul A este interior lui B.

2°. Oricare ar fi evenimentul A avem relaţiile :

$$\begin{aligned} A \cup A^* &= E \\ A \cup \Phi &= A \\ A \cup E &= E \\ E \cup \Phi &= E \end{aligned} \qquad (2_2)$$

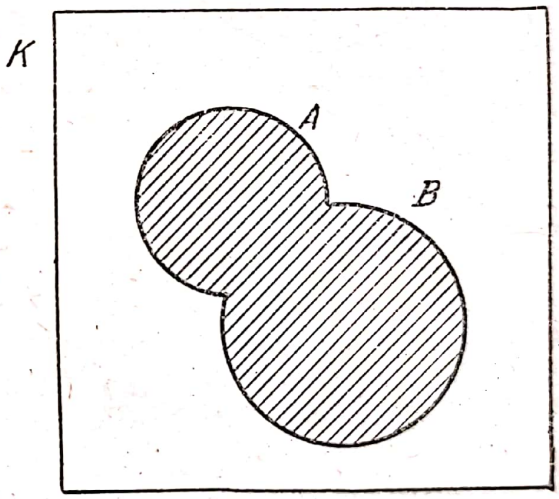

Fig. 13

3°. Dacă $A \subset C, B \subset C$, atunci $A \cup B \subset C$.

Intersecţia evenimentelor. A, B înseamnă realizarea simultană a celor două evenimente. Obţinem un nou eveniment P care se notează :

$$P = A \cap B \qquad (3)$$

şi se citeşte *A intersectat cu B*.

Geometric, $A \cap B$ este reprezentat prin regiunea comună celor două cercuri. În fig. 4, partea haşurată reprezintă evenimentul $A \cap B$.

Cu alte cuvinte evenimentul $A \cap B$ se realizează cînd un punct oarecare al lui K se găseşte în regiunea comună celor două cercuri.

În particular, oricare ar fi evenimentul A avem :

$$A \cap E = A \qquad (3_1)$$

Scanned with OKEN Scanner

căci regiunea comună pătratului şi cercului A este cercul A.

Se demonstrează uşor relaţiile :

$$\begin{aligned} A \cap \Phi &= \Phi \\ E \cap \Phi &= \Phi \\ A \cap A^* &= \Phi \end{aligned} \qquad (3_2)$$

Pentru aşezarea cercurilor din fig. 12 avem evident $A \cap B = \Phi$.

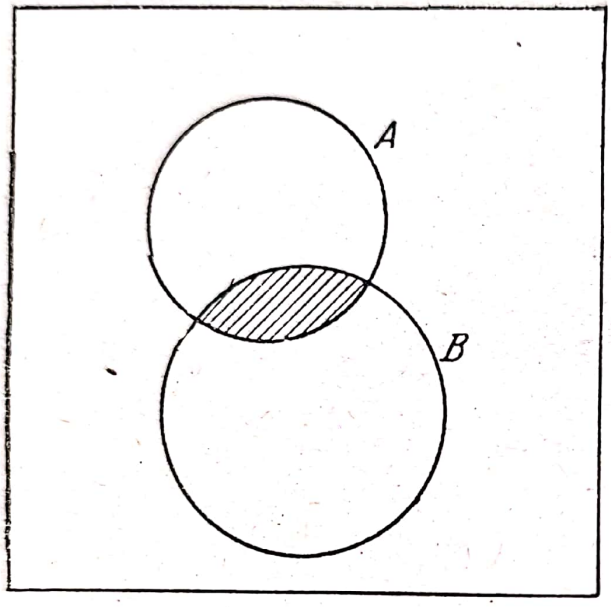

Fig. 14

Deoarece realizarea lui $A \cap B$ înseamnă realizarea simultană a evenimentelor A şi B rezultă implicaţiile

$$A \cap B \subset A, \qquad A \cap B \subset B \qquad (3_2)$$

Dacă $A \subset C$, $B \subset C$, atunci $A \cap B \subset C$.

Observaţie : $A \cap B^*$ este regiunea din cercul A pînă la cercul B (fig. 14), deoarece B^* înseamnă exteriorul lui B. Or, această regiune este compusă numai din puncte ale lui A care nu aparţin şi lui B.

Notînd-o prin $A - B$ avem relaţia :

$$A - B = A \cap B^* \qquad (4)$$

care defineşte *diferenţa evenimentelor A şi B*. Prin introducerea operaţiilor $\cup$ (*reuniune*) şi $\cap$ (*intersecţie*) unele noţiuni din teoria probabilităţilor pot fi formulate în mod mai precis. Aşa de pildă pentru evenimentele contrare avem definiţia : Evenimentele A şi A^* se numesc contrare dacă satisfac relaţiile :

$$A \cup A^* = E, \quad A \cap A^* = \Phi \qquad (5)$$

Evenimentele A şi B se zic *incompatibile între ele* dacă nu se pot realiza simultan, adică

$$A \cap B = \Phi \qquad (5_1)$$

În caz contrar, $(A \cap B \neq \Phi)$ evenimentele se zic *compatibile*.

Fie $\{A_1, A_2 \ldots, A_s\}$ o mulţime de evenimente pe care o putem nota mai simplu astfel

$$A = \{A_k\} \qquad 1 \leqslant k \leqslant s \qquad (6)$$

Vom spune că mulțimea (6) este o *desfacere în elemente incompatibile a evenimentului A*, dacă :

1°. $A_1 \cup A_2 \cup \ldots \cup A_s = A$

2°. $A_k \cap A_h = \Phi$ pentru orice k și h luînd fiecare valori de la 1 pînă la s, însă diferite între ele. Prima relație arată că A se realizează cînd se realizează sau A_1 sau A_2 sau ... sau A_s, indiferent care din ele, iar a doua că aceste evenimente sînt incompatibile între ele două cîte două.

Pentru ușurință condițiile 1° și 2° se scriu, de obicei, astfel :

$$1°.\ \bigcup_{k=1}^{s} A_k = A$$

$$2°.\ A_k \cap A_h = \Phi \qquad k,\ h = 1,\ 2,\ \ldots,s\ \ k \neq h \qquad (6_1)$$

Reuniunea a două evenimente se numește și *suma* lor, iar intersecția se numește și *produsul* evenimentelor. Scrierea $\bigcup_{k=1}^{s} A_k$ este analogă cu prescurtarea

$$A_1 + A_2 + \ldots + A_s = \sum_{k=1}^{s} A_k$$

În mod analog cu produsul prescurtat

$$A_1 \cdot A_2 \ldots A_s = \prod_{k=1}^{s} A_k$$

vom scrie adesea intersecția de evenimente :

$$A_1 \cap A_2 \cap \ldots \cap A_s = \bigcap_{k=1}^{s} A_k \qquad (6_2)$$

Operațiile $\cup$ și $\cap$ au proprietăți asemănătoare operațiilor de adunare și înmulțire cu numere :

$$1°.\quad A \cup B = B \cup A;\quad A \cap B = B \cap A \qquad (6_3)$$

(sînt comutative)

$$2°\quad A \cup (B \cup C) = (A \cup B) \cup C;\quad A \cap (B \cap C) = (A \cap B) \cap C \qquad (6_4)$$

(sînt asociative)

$$3°\quad A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \qquad (6_5)$$

(Intersecția este distributivă față de operația de reuniune).

Scanned with OKEN Scanner

Definiţie. O mulţime ordonată de evenimente se numeşte *cîmp de evenimente* dacă satisface următoarele axiome :
k_1 — Dacă evenimentul A aparţine cîmpului, atunci şi contrariul lui, A*, aparţine cîmpului. Notînd prin [E, K] cîmpul de evenimente, această axiomă se enunţă mai scurt. Dacă $A \in [E, K] \rightarrow A^* \in [E, K]$
Semnul $\rightarrow$ se va citi „rezultă".

k_2 — Dacă $A \in [E, K]$, $B \in [E, K] \rightarrow A \cup B \in [E, K]$
k_3 — Dacă $A \in [E, K]$, $B \in [E, K] \rightarrow A \cap B \in [E, K]$

Aceste trei axiome se numesc *axiomele cîmpului de evenimente* şi ele permit construirea sistematică a oricărui cîmp.

Corolarul I. Evenimentul sigur aparţine cîmpului

$$E \in [E, K] \tag{7}$$

În adevăr, dacă $A \in [E, K]$, după axioma k_1 şi $A^* \in [E, K]$, iar după axioma k_2 şi reuniunea $A \cup A^* = E \in [E, K]$

$$\Phi \in [E, K] \tag{7_1}$$

Corolarul II. După axioma k_1 şi corolarul precedent avem : $E \in [E, K] \rightarrow E^* = \Phi \in [E, K]$

Corolarul III. Dacă $A_1 \in [E, K]$, $A_2 \in [E, K]$ ... $A_r \in [E, K]$, atunci şi reuniunea lor

$$\bigcup_{k=1}^{r} A_k \in [E, K]$$

Corolarul IV. În mulţimea de evenimente a cîmpului [E, K] există o submulţime

$$\{ E_1, E_2, \ldots, E_n \} \tag{8}$$

numită *mulţimea evenimentelor elementare ale cîmpului*. Fiecare eveniment elementar E_k, $(1 \leqslant k \leqslant n)$ se bucură de proprietăţile :

1°. $E_k \neq \Phi$, pentru orice $k = 1, 2, \ldots n$

2°. Există cel puţin un eveniment $A \in [E, K]$ şi care nu este elementar aşa fel ca $E_k \subset A$]

Din această definiţie rezultă următoarele consecinţe :

1°. *Orice eveniment al cîmpului este o reuniune de evenimente elementare.* Notînd prin $i_1, i_2, \ldots i_r$ o combinaţie oarecare de r numere din şirul 1, 2, ..., n evenimentul

$$A_{i_1, i_2, i_3, \ldots i_r} = E_{i_1} \cup E_{i_2} \cup \ldots \cup E_{i_r} = \bigcup_{k=1}^{r} E_{i_k} \in [E, K] \tag{8_1}$$

după corolarul III.

Scanned with OKEN Scanner

Deoarece ordinea în care se reunesc evenimentele nu joacă un rol esențial putem presupune $1 \leqslant i_1 < i_2 < \ldots i_r \leqslant n$.

În particular, pentru $r = n$ avem *evenimentul sigur E, egal cu reuniunea evenimentelor elementare.*

$$E = \bigcup_{k=1}^{n} E_k \tag{2}$$

Pentru $r = 1$ avem mulțimea evenimentelor elementare

$$\{ E_k \} \qquad 1 \leqslant k \leqslant r$$

în număr de $n = C_n^1$

Pentru $r = 2$ avem mulțimea de evenimente

$$\left\{ A_{i_1, i_2} = \bigcup_{k=1}^{2} E_{i_k} \right\}$$

în număr de C_n^2.

Numărul evenimentelor cîmpului format din reuniuni de cîte r evenimente elementare este C_n^r.

În consecință, numărul evenimentelor cîmpului [E, K] este egal cu $\sum_{k=1}^{n} C_n^k$ la care adăugăm și evenimentul Φ, (în baza corolarului II), adică :

$$1 + \sum_{k=1}^{n} C_n^k = \sum_{k=0}^{n} C_n^k = 2^n$$

Această relație rezultă imediat din dezvoltarea

$$(1+1)^n = C_n^0 + C_n^1 + \ldots + C_n^n = \sum_{k=0}^{n} C_n^k$$

Cîmpul astfel construit se numește *cîmp finit de evenimente*, generat de mulțimea $\{E_1, E_2, \ldots E_n\}$ a evenimentelor elementare.

2°. *Evenimentele elementare sînt incompatibile între ele două cîte două*, adică $E_k \cap E_h = \Phi$ pentru orice k și $h = 1, 2, \ldots n$, $k \neq h$

În baza acestor relații și a formulei (8_2) deducem că mulțimea $\{E_k\}$ a evenimentelor elementare reprezintă o desfacere în elemente incompatibile a evenimentului sigur.

(O altă desfacere a evenimentului sigur este formată din mulțimea $\{A, A^*\}$ a două evenimente contrare).

Scanned with OKEN Scanner

Din cele expuse pînă acum rezultă modul de construire a cîmpului [E, K] :Vom fixa evenimentul sigur E şi componentele sale elementare $\{E_1, E_2, \ldots, E_n\}$. Combinînd două cîte două, evenimentele elementare, obţinem evenimentele $\{E_{i_1, i_2}\}$, trei cîte trei obţinem submulţimea $\{E_{i_1, i_2, i_3}\}$, . . . La acestea se adaugă evenimentul imposibil Φ.

Cîmpul are înfăţişarea de mai jos :

$$[E, K] = [\Phi ; \{E_k\} ; \{E_{i_1 i_2}\} ; \{E_{i_1 i_2 i_3}\}, \ldots \{E_{1, 2 \ldots n} = E\}] \qquad (9)$$

Aplicaţii. 1°. Să construim cîmpul de evenimente din exemplul 3°. Evenimentul sigur constă fie din extracţia becului de 60 lumini, fie din extracţia celui de 75 lumini. Acestea sînt evenimentele elementare pe care le notăm respectiv prin E_1 şi E_2, deci $n = 2$. Cîmpul va conţine $2^n = 2^2$ evenimente

$$[E,K] = \{\Phi ; E_1E_2 ; E = E_1 \cup E_2\}$$

2°. În exemplul 4° evenimentul sigur constă în apariţia oricăreia din cele 6 feţe, deci $n = 6$. Notînd prin E_k apariţia feţei cu K puncte $(1 \leqslant k \leqslant 6)$ avem :

$$[E,K] = [\Phi ; \{E_k\} ; \ldots ; E]$$

Cîmpul conţine $2^n = 2^6 = 64$ evenimente.

3°. În exemplul 5° evenimentul sigur se realizează o dată cu extracţia uneia oarecare din cele 17 piese. Deci $n = 17$. Cîmpul va avea 2^{17} evenimente.

4°. Să presupunem că fiecare probă constă din extracţia a două piese din lot. Evenimentul sigur este extracţia uneia oarecare din perechile care se pot forma din cele 17 piese. În acest caz $n = C_{17}^2$, deci cîmpul va avea $2\ C_{17}^2$ evenimente.

Probleme

1. Dacă A şi B aparţin cîmpului [E, K] să se arate :

$$1°.\ (A \cup B)^* = A^* \cap B^*$$

$$2°.\ (A \cap B)^* = A^* \cup B^*$$

2. Să se demonstreze că relaţiile: $A \subset B$, $B^* \subset A^*$, $A \cup B = B$, $A \cap B = A$ sînt echivalente.

3. Relaţiile $A \cup B$, $A^* \supset B$, $B^* \subset A$ sînt echivalente.

4. Dacă evenimentele A, B, C aparţin aceluiaşi cîmp [E,K] avem :

$$1°.\ (A \cup B \cup C)^* = A^* \cap B^* \cap C^*$$

$$2°.\ (A \cup B)^* \cap C = (A \cap C)^* \cup (B^* \cap C)$$

Scanned with OKEN Scanner

5. Să se verifice relația $A \cup B = A \cup (B - A \cap B)$

6. Noțiunile de reuniune și intersecție se extind la orice mulțime infinită, numărabilă sau nu de evenimente.

Fie $A_1, A_2 .. A_n, \ldots = \{ A_k \}_{k \in N}$ o mulțime numărabilă de evenimente; N însemnează mulțimea numerelor naturale. Reuniunea: $S = \bigcup_{k=1}^{\infty} A_k$ constă în aceea că se realizează cel puțin unul din evenimentele A_k, iar intersecția $P = \bigcap_{k=1}^{\infty} A_k$ însemnează realizarea simultană a tuturor evenimentelor. Sumele $S_n = \bigcup_{k=1}^{n} A_k$ și produsele $P_n = \bigcap_{k=1}^{n} A_k$ se numesc respectiv sume și produse parțiale de ordinul n. Avem relațiile evidente:

$$S_1 \supset S_2 \subset \ldots$$

$$P_1 \subset P_2 \subset \ldots$$

$$S = \bigcup_{n=1}^{\infty} S_n, \quad P = \bigcap_{n=1}^{\infty} P_n$$

Să se demonstreze relațiile:

a) $\left\{ \bigcup_{k=1}^{\infty} A_k \right\} \bigcup \left\{ \bigcap_{k=1}^{\infty} A_k^* \right\} = E$

b) $\left\{ \bigcup_{k=1}^{\infty} A_k \right\} \bigcap \left\{ \bigcap_{k=1}^{\infty} A_k^* \right\} = \Phi$

c) $\bigcap_{k=1}^{\infty} A_k^* = \left(\bigcup_{k=1}^{\infty} A_k \right)^*$

d) $\bigcup_{k=1}^{\infty} A_k = \left(\bigcap_{k=1}^{\infty} A_k^* \right)^*$

Scanned with OKEN Scanner

Capitolul III

ROBABILITĂŢI

§1. Definiţia axiomatică a probabilităţii

Să considerăm exemplul 5° din §1 Cap. II şi să notăm prin A_1 evenimentul de a extrage o piesă stan ard şi A_2 evenimentul de a extrage o piesă rebut. Este evident că această situaţie şi altele asemănătoare impune alegerea unui anumit mod de a măsura şansele de realizare ale celor două evenimente.

D e f i n i ţ i e. Numărul P (A) se numeşte probabilitatea evenimentul $A \in [E, K]$, definită pe acest cîmp, dacă satisface axiomele :

P_1 — $P(A) \geqslant 0$, pentru orice $A \in [E, K]$

P_2 — $P(E) = 1$

P_3 — Dacă $A \in [E, K]$, $B \in [E, K]$ şi dacă $A \cap B = \Phi$ atunci : $P(A \cup B) = P(A) + P(B)$

Aceste trei axiome se numesc *axiomele probabilităţii*. Axioma P_1 afirmă că probabilitatea oricărui eveniment al cîmpului este un număr *nenegativ*. În baza axiomei P_2, evenimentul sigur are probabilitatea egală cu unu, iar axioma P_3 ne permite să calculăm probabilitatea reuniunii a două evenimente incompatibile între ele cînd sînt cunoscute probabilităţile fiecăruia din evenimente. Această a treia axiomă se numeşte şi *regula de adunare a probabilităţilor*. Condiţia de incompatibilitate este esenţială deoarece s-ar putea ca suma P (A) + P (B) să fie mai mare ca unu cînd evenimentele nu sînt incompatibile. Un asemenea rezultat este absurd, deoarece vom arăta mai departe că *probabilitatea oricărui eveniment este cel puţin egală cu zero şi cel mult egală cu unu*. Din axiomele $P_1 - P_3$ decurg unele consecinţe utile.

Scanned with OKEN Scanner

Corolarul I. Probabilitatea evenimentului imposibil este nulă:

$$P(\Phi) = 0 \tag{1}$$

In adevăr $\Phi \cup E = E$ şi cum $\Phi \cap E = \Phi$ după axioma P_3 avem $P(\Phi \cup E) = P(\Phi) + P(E) = P(E)$. De unde $P(\Phi) = 0$.

Corolarul II. Am văzut în capitolul II că pentru orice eveniment al cîmpului $A \in [E, K]$ avem desfacerea:

$$A = \bigcup_{k=1}^{s} E_{i_k} \qquad (s < n)$$

unde $E_k \in [E, K]$ şi $E_k \cap E_h = \Phi$, $k,h, = 1, 2, \ldots n$; $h \neq k$, deoarece sînt evenimente elementare. Pe baza axiomei P_3 putem scrie relaţia mai generală

$$P\left(\bigcup_{k=1}^{s} E_{i_k}\right) = \sum_{k=1}^{s} P(E_{i_k}) \tag{1_1}$$

Vom demonstra-o prin recurenţă. Pentru $s = 2$ este adevărată în baza axiomei P_3. Admitem

$$P\left(\bigcup_{k=1}^{s-1} E_{i_k}\right) = \sum_{k=1}^{s-1} P(E_{i_k})$$

şi să demonstrăm valabilitatea relaţiei (1_1)

$$\text{Notăm: } \bigcup_{k=1}^{s-1} E_{i_k} = A, \text{ atunci } \bigcup_{k=1}^{s} E_{i_k} = A \cup E_{i_s}.$$

Deoarece $A \cap E_{i_s} = \Phi$, căci fiecare element al lui A este incompatibil cu E_{i_s}, fiind evenimente elementare, avem după P_3:

$$P(A \cup E_{i_s}) = P(A) + P(E_{i_s}) = \sum_{k=1}^{s-1} P(E_{i_k}) + P(E_{i_s})$$

adică ceea ce trebuia să demonstrăm.

Corolarul III. Pentru $s = n$ avem:

$$P\left(\bigcup_{k=1}^{n} E_k\right) = P(E) = 1$$

$$\text{Dar } P\left(\bigcup_{k=1}^{n} E_k\right) = \sum_{k=1}^{n} P(E_k) = \sum_{k=1}^{n} P_k$$

Scanned with OKEN Scanner

unde am notat $P(E_k) = P_k$. Rezultă deci :

$$P_1 + P_2 + \ldots + P_n = 1 \qquad (1_2)$$

adică *suma probabilităților evenimentelor elementare este egală cu 1.*

În particular, pentru evenimentele contrare, axioma P_3 ne conduce la relația

$$P(A) + P(A^*) = 1 \qquad (1_3)$$

Corolarul IV. Dacă $A \in [E, K]$, $B \in [E, K]$ și dacă $A \subset B$ atunci $P(A) \leqslant P(B)$.

În adevăr A și B sînt de forma

$$A = \bigcup_{k=1}^{r} E_k, \qquad B = \bigcup_{k=1}^{r+s} E_k \quad (r + s \leqslant n)$$

De unde deducem :

$$P(A) = \sum_{k=1}^{r} P_k, \quad P(B) = \sum_{k=1}^{r+s} P_k = \sum_{k=1}^{r} P_k + \sum_{r+1}^{r+s} P_k = P(A) + \sum_{k=r+1}^{r+s} P_k$$

După axioma P_1 suma $\sum_{k=r+1}^{r+s} P_k \geqslant 0$ deci $P(B) \geqslant P(A)$

Corolarul V. Pentru orice $A \in [E, K]$ avem :

$$0 \leqslant P(A) \leqslant 1 \qquad (1_4)$$

Deoarece $\Phi \cup A = A$ rezultă $\Phi \subset A$ deci, după corolarul IV, avem : $P(\Phi) \leqslant P(A)$. Și de asemenea, $A = A \cap E \subset E$ de unde $P(A) \leqslant P(E)$. Așadar, am găsit $P(\Phi) \leqslant P(A) \leqslant P(E)$ sau încă $0 \leqslant P(A) \leqslant 1$

§ 2. Definiția clasică a probabilității

Fie $[E, K]$ un cîmp finit de evenimente generat de mulțimea $\{E_k\}$, $(1 \leqslant k \leqslant n)$ a evenimentelor sale elementare și

$$E_{i_1, i_2, \ldots i_s} = \bigcup_{k=1}^{s} E_{i_k}$$

un eveniment din acest cîmp. Am văzut mai înainte că :

$$P(E_{i_1, i_2, \ldots i_s}) = P(E_{i_1}) + P(E_{i_2}) + \ldots + P(E_{i_s})$$

Deoarece evenimentele elementare au aceeași șansă de a se realiza, în fiecare probă este firesc să admitem :

$$P(E_{i_1}) = P(E_{i_2}) = \ldots = P(E_{i_n})$$

Din relația (1_2) deducem atunci :

$$P_k = P(E_k) = \frac{1}{n} \qquad (2)$$

pentru orice $k = 1, 2, \ldots, n$. Evenimentul $E_{i_1, i_2, \ldots i_s}$ fiind reuniunea a s evenimente elementare, în baza relației (1_1) va avea probabilitatea

$$P(E_{i_1, i_2, \ldots, i_s}) = \frac{s}{n} \qquad (2_1)$$

Exemple : 1°. Să notăm prin A extracția unei piese standard dintr-un lot de 17 piese din care 15 piese sînt standard. Vom avea : $n = 17$, $s = 15$, deci $P(A) = \frac{s}{n} = \frac{15}{17}$

2°. În cazul aruncării unui zar avem $n = 6$. Deci probabilitatea apariției unei fețe anumite ($s = 1$) va fi egală cu $\frac{1}{6}$.

Dacă evenimentul A constă din apariția unui număr fără soț avem : $A = E_1 \cup E_3 \cup E_5$ de unde $s = 3$. Atunci : $P(A) = \frac{3}{6} = \frac{1}{2}$.

3°. Presupunînd că fiecare probă constă din extracția unei perechi de piese din cele 17 din exemplul 1°, probabilitatea ca să obținem o pereche formată dintr-o piesă standard și una rebut se va calcula în modul următor :

După cum se știe, numărul evenimentelor elementare este $n = C_{17}^2$. Din acestea să calculăm numărul celor care răspund problemei, adică perechea extrasă să conțină piese de calități diferite. Este evident că o piesă din cele 15 piese standard poate fi asociată cu fiecare din cele două piese rebut și obținem perechi de piese care satisfac enunțul ; la cele 15 piese corespund deci 15×2 perechi de astfel de piese. Așadar, din cele $n = C_{17}^2$ evenimente, $s = 15 \times 2$ răspund problemei puse, deci probabilitatea căutată va fi egală cu $\frac{15 \times 2}{C_{17}^2}$.

Mai general, putem rezolva următoarea problemă. Într-o urnă sînt a bile albe și b bile negre. Se extrag deodată n bile (sau ceea ce este tot una se extrag succesiv n bile fără a pune bila înapoi în urnă). Să se calculeze probabilitatea ca α din ele să

Scanned with OKEN Scanner

fie albe şi β ($= n - \alpha$) să fie negre. Notînd prin A acest eveniment, probabilitatea căutată se calculează cu formula

$$P(A) = \frac{C_a^{\alpha} \cdot C_b^{\beta}}{C_{a+b}^{\alpha+\beta}} \qquad (2_2)$$

În adevăr, numărul evenimentelor elementare este egal cu numărul grupelor de $n = \alpha + \beta$ elemente din cele $a + b$, indiferent de ordinea elementelor în grupă, adică $C_{a+b}^{\alpha+\beta}$. Din acestea să evaluăm numărul acelor evenimente elementare care răspund enunţului, adică α bile să fie albe şi β negre. Evident, din cele a bile albe putem forma C_a^{α} grupe de α bile albe. Fiecare grupă din acestea poate fi asociată cu fiecare din cele C_b^{β} grupe de β bile negre şi obţinem grupe de $\alpha + \beta = n$ bile din care α sînt albe şi β negre. Numărul acestora, deci al evenimentelor elementare, va fi egal cu $C_a^{\alpha} \cdot C_b^{\beta}$. Probabilitatea căutată este deci (2_2).

Problema încă se poate extinde presupunînd că în urnă sînt a_1 bile de culoarea 1, a_2 bile de culoarea 2,... a_s bile de culoarea s. Se extrag deodată n bile şi se cere probabilitatea ca α_1 să fie de culoarea 1,... α_s de culoarea s. $\left(\sum_{k=1}^{s} \alpha_k = n\right)$.

Vom avea :

$$P(A) = \frac{C_{a_1}^{\alpha_1} \cdot C_{a_2}^{\alpha_2} \cdots C_{a_s}^{\alpha_s}}{C_{a_1+a_2+\ldots+a_s}^{\alpha_1+\alpha_2+\ldots+\alpha_s}} \qquad (2_3)$$

Această problemă, cunoscută sub numele de „*schema bilei nereverite*" are importante aplicaţii în teoria selecţiei.

Cîmp finit de probabilitate. Am văzut că fiecărui eveniment $A \in [E,K]$ îi corespunde în mod univoc numărul P (A), probabilitatea sa. Mulţimea valorilor P (A), împreună cu evenimentele cîmpului [E,K] formează un *cîmp finit de probabilitate* pe care îl notăm prin [E, K, P].

§ 3. Probabilităţi condiţionate

Definiţie. Funcţia

$$P_A(B) = \frac{P(A \cap B)}{P(A)} \qquad (3)$$

unde $P(A) \neq 0$ se numește *probabilitatea evenimentului B condiționată de realizarea evenimentului A.*

Uneori, probabilitatea condiționată $P_A(B)$ se mai notează și prin $P(B/A)$ așa încît,

$$P(B/A) = \frac{P(A \cap B)}{P(A)} \qquad (3')$$

Pentru ca definiția de mai sus să aibă sens trebuie să arătăm că probabilitatea $P_A(B)$ verifică axiomele $P_1 - P_3$.

Deoarece $P(A \cap B)$ și $P(A)$ sînt nenegative (axioma P_1) rezultă $P_A(B) \geqslant 0$. Pentru verificarea axiomei P_2 să demonstrăm relația $P_A(E) = 1$. În acest scop, vom face în (3) $B = E$ și ținînd seama că $A \cap E = A$ găsim :

$$P_A(E) = \frac{P(A \cap E)}{P(A)} = \frac{P(A)}{P(A)} = 1$$

În ce privește axioma P_3, să considerăm evenimentele B și C incompatibile între ele și să demonstrăm relația $P_A(B \cup C) = P_A(B) + P_A(C)$

După definiția (3) putem scrie

$$P_A(B \cup C) = \frac{P[A \cap (B \cup C)]}{P(A)} = \frac{P[(A \cap B) \cup (A \cap C)]}{P(A)}$$

Dar evenimentele $A \cap B$ și $A \cap C$ sînt incompatibile căci B și C sînt incompatibile prin ipoteză ; atunci după axioma P_3 putem scrie :

$P[(A \cap B) \cup (A \cap C)] = P(A \cap B) + P(A \cap C)$ Relația precedentă devine :

$$P_A(B \cup C) = \frac{P(A \cap B) + P(A \cap C)}{P(A)}$$

În baza relației (3), însă avem : $P(A \cap B) = P(A) \cdot P_A(B)$; $P(A \cap C) = P(A) \cdot P_A(C)$. Înlocuind aceste valori în $P_A(B \cup C)$ obținem tocmai relația ce trebuia dovedită. Prin urmare, mulțimea $[E, K, P_A]$ este un cîmp finit de probabilitate.

C o r o l a r u l I. Probabilitatea lui A condiționată de realizarea lui B, după definiția (3) va fi :

$$P_B(A) = \frac{P(A \cap B)}{P(B)} \qquad (3'')$$

Din (3) și (3″) deducem :

$$P(A \cap B) = P(A) \cdot P_A(B) = P(B) \cdot P_B(A) \qquad (3_1)$$

C o r o l a r u l II. Relația $P(A \cap B) = P(A) \cdot P_A(B)$ se poate generaliza pentru s evenimente și anume :

$$P\left(\bigcap_{k=1}^{s} A_k\right) = P(A_1) \cdot P_{A_1}(A_2) \cdot P_{A_1 \cap A_2}(A_3) \cdot \ldots \cdot P\left(A_s \Big/ \bigcap_{k=1}^{s-1} A_k\right) \qquad (3_2)$$

sau folosind relația (3′) : $P(A_1 \cap A_2 \cap \ldots \cap A_s) = P(A_1) \cdot P(A_2/A_1) \cdot P(A_3/A_1 \cap A_2) \ldots P(A_s/A_1 \cap A_2 \cap A_{s-1})$

Această relație se demonstrează prin recurență :

Pentru $s = 2$ avem relația (3_1). Admitem că relația este devărată pentru $(s - 1)$ evenimente, adică :

$$P\left(\bigcap_{k=1}^{s-1} A_k\right) = P(A_1) \cdot P_{A_1}(A_2) \ldots P_{\bigcap_{k=1}^{s-2} A_k}(A_{s-1})$$

și să dovedim relația (3_2). Notînd $A = \bigcap_{k=1}^{s-1} A_k$ avem relația

$$P(A \cap A_s) = P(A) \cdot P_A(A_s)$$

valabilă după (3_1) pentru evenimentele A și A_s. Deoarece putem scrie :

$$A \cap A_s = \left(\bigcap_{k=1}^{s-1} A_k\right) \cap A_s = \bigcap_{k=1}^{s} A_k$$

vom găsi :

$$P\left(\bigcap_{k=1}^{s} A_k\right) = P(A_1) \cdot P_{A_1}(A_2) \ldots P_{\bigcap_{k=1}^{s-2} A_k}(A_{s-1}) \cdot P_{\bigcap_{k=1}^{s-1} A_k}(A_s)$$

deci relația ce trebuia demonstrată.

Formula (3_2) constituie *regula de înmulțire a probabilităților*, care împreună cu regula de adunare, aplicate simultan sau separat ne ajută să calculăm probabilitatea unor evenimente destul de complicate.

E v e n i m e n t e i n d e p e n d e n t e. Prin definiție, evenimentul B este independent de evenimentul A dacă verifică relația :

$$P_A(B) = P(B) \qquad (3_3)$$

Corolar. Înlocuind în (3_1) obținem :

$$P(A) \cdot P(B) = P(B) \cdot P_B(A)$$

adică : $P(A) = P_B(A)$ (3_4)

Așadar, *dacă B este independent de A, atunci reciproc și A este independent de B.*

Evenimentele $A_1, A_2, \ldots, A_s$ se zic *independente în totalitatea lor*, dacă pentru orice submulțime $\{A_{i_1}, A_{i_2}, \ldots A_{i_r}\}$ unde $1 \leqslant i_1 \leqslant \ldots \leqslant i_r \leqslant s$ avem :

$$P\left(\bigcap_{k=1}^{r} A_{i_k}\right) = P(A_{i_1}) \cdot P(A_{i_2}) \ldots P(A_{i_r}) \qquad (3_5)$$

Din această definiție rezultă că pentru independența globală este necesar ca evenimentele elementare să fie independente două cîte două, adică : $P(E_{i_k} \cap E_{i_h}) = P(E_{i_k}) \cdot P(E_{i_h})$ pentru orice $k, h = 1,2,\ldots r$; $k \neq h$. Această condiție nu este însă și suficientă după cum rezultă din următorul exemplu : Fie mulțimea de evenimente elementare $\{E_k\}$, $1 \leqslant k \leqslant 4$ și $A_1 = E_1 \cup E_2$, $A_2 = E_1 \cup E_3$, $A_3 = E_1 \cup E_4$
Vom avea probabilitățile :

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{2}$$

$$P(A_1 \cap A_2) = P(A_1 \cap A_3) = P(A_2 \cap A_3) = \frac{1}{4} = \left(\frac{1}{2}\right)^2$$

$$P(A_1 \cap A_2 \cap A_3) = \frac{1}{4} \neq \left(\frac{1}{2}\right)^3$$

Dacă evenimentele sînt independente relația (3_2) se scrie mai simplu :

$$P\left(\bigcap_{k=1}^{s} A_k\right) = P(A_1) \cdot P(A_2) \ldots P(A_s) \qquad (3_2'')$$

Exemple : 1°. Evenimentul A are aceeași șansă de a se realiza în fiecare din cele n probe care se fac independent una de alta. Să se determine probabilitatea ca în aceste n probe evenimentul A să se realizeze de x ori și contrariul său A^* de n-x ori.

Acest gen de probleme se încadrează în următoarea schemă teoretică : o urnă conține a bile albe și b bile de alte culori. Se fac n extracții punînd bila înapoi după fiecare extracție. Notînd prin A evenimentul de a extrage o bilă albă din urnă și prin A^* contrariul său, adică de a extrage o bilă de altă culoare decît

Scanned with OKEN Scanner

cea albă, se cere probabilitatea ca A să se producă de x ori în n probe şi A* de n-x ori.

Această schemă se numeşte *schema bilei revenite* sau *schema lui Bernoulli*. Să notăm pentru prescurtarea scrisului $P(A) = p$, atunci $P(A^*) = 1 - p = q$

Evenimentul care ne interesează (din n probe să obţinem A de x ori şi A* de n-x ori) se poate realiza în mai multe moduri, fie ele $A_1, A_2 \ldots, A_s$ incompatibile două cîte două, evident. Numărul acestor moduri s este egal cu numărul permutărilor de n elemente din care x sînt egale cu A şi n-x egale cu A*. Aşadar, s este egal cu numărul permutărilor cu repetiţie

$$s = P_n^{(x,\, n-x)} = \frac{n!}{x!\,(n-x)!} \qquad (3_6)$$

Probabilitatea să se producă A_k, de pildă, $(1 \leqslant k \leqslant s)$ se calculează cu formula $(3'_2)$ şi este egală cu un produs de x factori egali cu p şi (n—x) factori egali cu q, deci $P(A_k) = p^x q^{n-x}$ pentru orice $k = 1, 2, \ldots, s$. Deoarece $A_1, A_2, \ldots, A_s$ sînt incompatibile, în baza regulii de adunare avem : $P(A) = P(A_1) + P(A_2) + \ldots + P(A_s) = s \cdot p^x q^{n-x}$

sau :

$$P(A) = \frac{n!}{x!\,(n-x)!}\, p^x q^{n-x} \qquad (3_7)$$

Mai general, dacă o urnă conţine a_1 bile de culoarea $1, \ldots, a_s$ bile de culoarea s şi se fac n extracţii punînd de fiecare dată bila înapoi în urnă, probabilitatea ca să extragem x_1 bile de culoarea $1, \ldots x_s$ bile de culoarea s se calculează cu formula

$$P(A) = \frac{n!}{x_1!\, x_2! \ldots x_s!}\, p_1^{x_1} \cdot p_2^{x_2} \ldots p_1^{x_s} \qquad (3'_7)$$

unde $p_k = \dfrac{a_k}{\sum_{k=1}^{s} a_k}$ pentru orice $k = 1,2, \ldots s$ şi $\sum_{k=1}^{s} x_k = n$.

Această formulă constituie o generalizare a formulei (3_7) şi se demonstrează analog.

2°. Dintr-un lot de 17 piese din care 15 sînt piese standard se extrag succesiv două piese, fără a pune piesa extrasă înapoi în lot. Se cere probabilitatea ca aceste două piese să fie de calităţi diferite.

Acest eveniment A, se poate realiza în două moduri incompatibile :

a) În prima extracție obținem o piesă standard, iar în a doua o piesă rebut, sau

b) În prima extracție o piesă rebut, iar în a doua o piesă standard.

Notînd A_1 și A_2 evenimentele a) și b) respectiv, avem $A = A_1 \cup A_2$, $A_1 \cap A_2 = \Phi$

$$P(A) = P(A_1) + P(A_2) \qquad \text{(axioma } P_3\text{)}$$

Dacă A_1' constă în extracția piesei standard și A_2' este extracția unei piese rebut avem :

$$P(A_1) = P(A_1') \cdot P_{A_1'}(A_2')$$
$$P(A_2) = P(A_2') \cdot P_{A_2'}(A_1')$$

Deoarece $P(A_1') \cdot P_{A_1'}(A_2') = P(A_2') \cdot P_{A_2'}(A_1')$ (după relația 3_1) putem scrie : $P(A) = 2\ P(A_1') \cdot P_{A_1'}(A_2') = 2 \cdot \frac{15}{17} \cdot \frac{2}{16}$ deoarece $P(A_1') = \frac{15}{17}$, iar $P_{A_1'}(A_2') = \frac{2}{16}$. Întrucît cele două piese rebut le extragem acum dintr-un lot de 16 piese. Putem încă scrie

$$P(A) = \frac{15 \times 2}{\left(\frac{17 \times 16}{2}\right)} = \frac{15 \cdot 2}{C_{17}^2}$$

adică rezultatul găsit mai înainte pe altă cale.

Regula lui Bayes. Fie $A_1, A_2, \ldots A_s$ o desfacere în elemente incompatibile a evenimentului sigur. În baza definiției avem :

$$\bigcup_{k=1}^{s} A_k = E,\ A_k \cap A_h = \Phi;\ k, h = 1, 2, \ldots s,\ k \neq h$$

Pentru un eveniment oarecare $B \in [E, K, P]$ avem desfacerea :

$$B = \bigcup_{k=1}^{s} (A_k \cap B) \qquad (3_4)$$

În adevăr, în relația $\bigcup_{k=1}^{s} A_k = E$ intersectînd cu B obținem :

$$B \cap \left(\bigcup_{k=1}^{s} A_k \right) = E \cap B$$

Scanned with OKEN Scanner

sau încă :

$$\bigcup_{k=1}^{s} (A_k \cap B) = B$$

De altă parte avem relația : $P(A_k) \cdot P_{A_k}(B) = P(B) \cdot P_B(A_k)$

de unde :

$$P_B(A_k) = \frac{P(A_k) \cdot P_{A_k}(B)}{P(B)}$$

Dar :

$$P(B) = P\left[\bigcup_{k=1}^{s} (A_k \cap B)\right] = \sum_{k=1}^{s} P(A_k \cap B) = \sum_{k=1}^{s} P(A_k) \cdot P_{A_k}(B)$$

Înlocuind mai sus, obținem regula lui Bayes :

$$P_B(A_k) = \frac{P(A_k) \cdot P_{A_k}(B)}{\sum_{k=1}^{s} P(A_k) \cdot P_{A_k}(B)} \qquad (4)$$

Exemplu : La un magazin sosesc două lăzi identice ca formă, cu pantofi de copii ; una de la fabrica A și cealaltă de la fabrica B. Din pantofii produși de fabrica A, se știe că 1% au defecte, iar cei de la fabrica B,2% au defecte. Din una din lăzi, la întîmplare se ia o pereche de pantofi și se constată că au defecte. Să se determine probabilitatea ca această pereche să provină de la fabrica B.

Notăm prin A_1 extracția din lada de la fabrica A și prin A_2 extracția din lada de la fabrica B. Fie x evenimentul ca o pereche de pantofi să aibă defecte. După formula (4) avem :

$$P_z(A_2) = \frac{P(A_2) \cdot P_{A_2}(x)}{P(A_1) \cdot P_{A_1}(x) + P(A_2) \cdot P_{A_2}(x)}$$

din datele problemei avem însă :

$$P(A_1) = P(A_2) = \frac{1}{2}$$

$$P_{A_1}(x) = \frac{1}{100} \cdot P_{A_2}(x) = \frac{2}{100} \text{ deci}$$

$$P_X(A_2) = \frac{\frac{1}{2} \cdot \frac{2}{100}}{\frac{1}{2} \cdot \frac{1}{100} + \frac{1}{2} \cdot \frac{2}{200}} = \frac{2}{3}.$$

Scanned with OKEN Scanner

§4. Evenimente compatibile

Evenimentele A şi B sînt compatibile dacă se pot realiza în acelaşi timp ; în scris $A \cap B \neq \Phi$.

Să considerăm figura de mai jos :

Partea simplu haşurată reprezintă evenimentul A_1, iar cea dublu haşurată intersecţia exteriorului lui A_1 (adică, A_1) cu evenimentul A_2. Cum partea simplu şi dublu haşurată este reuniunea $A_1 \cup A_2$ putem scrie relaţia

$$A_1 \cup A_2 = A_1 \cup (A_1^* \cap A_2)$$

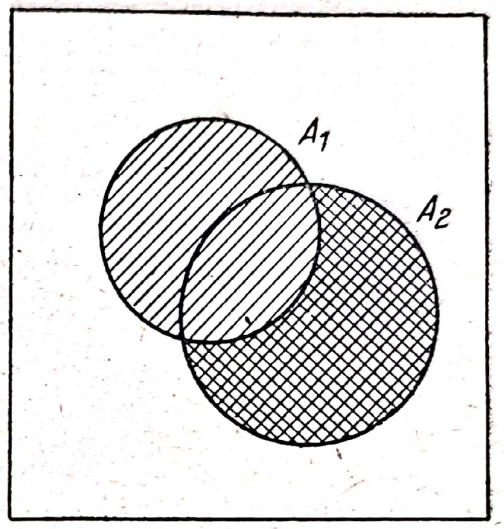

Fig. 15

Evenimentele A_1 şi $A_1^* \cap A_2$ sînt incompatibile, deci :

$$P(A_1 \cup A_2) =$$
$$= P(A_1) + P(A_1^* \cap A_2) =$$
$$= P(A_1) + P[(E - A_1 \cap A_2] =$$
$$= P(A_1) + P(A_2 - A_1) \cap A_2).$$

Pentru calculul ultimei probabilităţi vom demonstra următoarea *lemă*. Fie α şi β două evenimente din acelaşi cîmp. Dacă $\alpha \leqslant \beta$, atunci : $P(\beta - \alpha) = P(\beta) - P(\alpha)$

În adevăr, avem identitatea : $\beta = \alpha \cup (\alpha^* \cap \beta)$

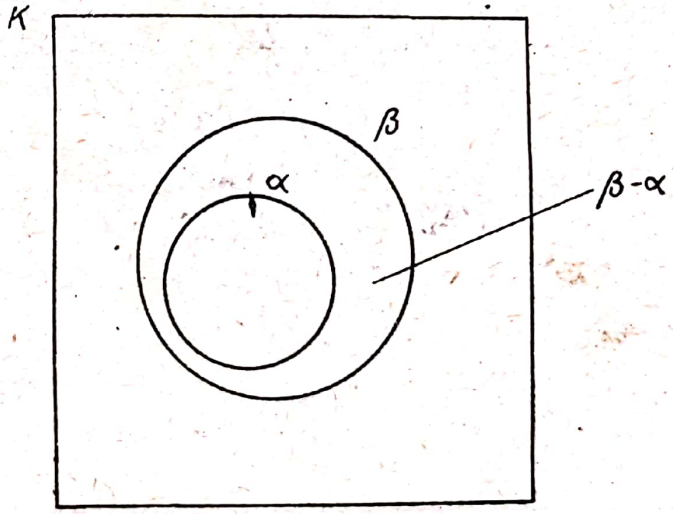

Fig. 16

Cum $\alpha \cap (\alpha^* \cap \beta) = \Phi$, putem scrie : $P(\beta) = P(\alpha) + P(\alpha^* \cap \beta)$

Dar $\alpha^* \cap \beta = \beta - \alpha$ (relaţia 4 Capitolul II) deci : $P(\beta) = P(\alpha) + P(\beta - \alpha)$

În baza acestei teme obținem :

$$P(A_1 \cup A_2) = P(A_1) + P(A_2 - A_1 \cap A_2) =$$
$$= P(A_1) + P(A_2) - P(A_1 \cap A_2) \qquad (5)$$

întrucît $A_1 \cap A_2 \subset A_2$

Această relație se poate generaliza pentru s evenimente. Prin recurență se obține formula lui Poincaré

$$P\left(\bigcup_{k=1}^{s} A_k\right) = \sum_{k=1}^{s} P(A_k) - \sum_{\substack{k,\ h \\ k<h}} P(A_k \cap A_h) + \ldots + (-1)^{s-1} P\left(\bigcap_{k=1}^{s} A_k\right) \qquad (5')$$

Această formulă dă *probabilitatea ca cel puțin unul din cele s evenimente* $A_1, A_2, \ldots A_s$, *să se întîmple.*

În particular, dacă evenimentele A_k sînt incompatibile două cîte două ($A_k \cap A_h = \Phi$), obținem relația cunoscută :

$$P\left(\bigcup_{k=1}^{s} A_k\right) = \sum_{k=1}^{s} P(A_k) \qquad (5'')$$

Exemplu. La un atelier de montare piesele sosesc în loturi de cîte 5 aranjate în ordinea 1, 2, 3, 4, 5. Printr-un accident în timpul transportului la atelier, piesele dintr-un lot au fost amestecate. Să se determine probabilitatea ca cel puțin o piesă din acest lot să fie montată în ordinea ei normală.

Să notăm prin A_k evenimentul ca piesa cu numărul k să se găsească la locul marcat k în lot ($1 \leqslant k \leqslant 5$) și să evaluăm probabilitățile $P(A_k)$, $P(A_k \cap A_h) \ldots P(A_1 \cap A_2 \cap \ldots \cap A_5)$ care intră în formula (5'). Evenimentele A_1, A_2 A_3, A_4, A_5 sînt compatibile evident, căci se poate ca mai multe din cele 5 piese să se găsească pe locul lor normal.

Să calculăm, de pildă, $P(A_3)$. Dacă piesa cu numărul 3 se găsește pe locul al treilea celelalte 4 pot forma P_4 grupe de 4 piese. Toate cele 5 piese formează în total P_5 grupe. O atare grupă este un eveniment elementar, iar din acestea P_4 răspund problemei. Avem deci probabilitatea $P(A_3) = \frac{P_4}{P_5} = \frac{1}{5}$ și în general, $P(A_k) = \frac{1}{5}$, $1 \leqslant k \geqslant 5$.

În mod analog obținem

$$P(A_k \cap A_h) = \frac{1}{5 \cdot 4} \qquad k, h = 1, 2, \ldots, 5, (k < h)$$

. . . . . .

$$P\left(\bigcap_{k=1}^{5} A_k\right) = \frac{1}{5!}$$

Evenimentele A_k sînt în număr de 5 ; $A_k \cap A_h$ sînt în număr de C_5^2 etc., deci înlocuind în (5′) găsim :

$$P\left(\bigcup_{k=1}^{5} A_k\right) = 1 - C_5^2 \cdot \frac{1}{5 \cdot 4} + C_5^3 \frac{1}{5 \cdot 4 \cdot 3} - C_5^4 \frac{1}{5 \cdot 4 \cdot 3 \cdot 2} + \frac{1}{5!} =$$

$$= 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \frac{1}{5!}$$

Probabilitatea ca nici o piesă să nu fie montată în ordinea sa normală este evident

$$1 - P\left(\bigcup_{k=1}^{5} A_k\right) = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!}$$

deoarece evenimentele „cel puțin una” și „nici una” sînt contrare unul altuia.

Probleme

1. Se execută angrenaje de cîte două roți dințate una avînd diametrul standard 1,5 m și cealaltă 0,9 m. La atelier sosesc două loturi de cîte 15 roți respectiv din prima și a doua categorie. Din primul lot 3 roți au diametrul sub 1,5 m și 2 peste 1,5 m, iar din al doilea 5 au diametrul sub 0,9 m și una peste 0,9 m. Angrenajul este „bun” dacă roțile au dimensiunile standard. Să se calculeze probabilitatea ca un angrenaj să fie bun.

$$R : \frac{90}{225}$$

2. Dintr-un lot de 34 aparate telefonice 6 nu fac apel. Se iau la întîmplare 12 aparate din acest lot și se cere probabilitatea ca 2 din ele să nu facă apel.

$$R : \frac{1}{C_{34}^6} C_6^2 \cdot C_{28}^4$$

Scanned with OKEN Scanner

3. În problema precedentă să se calculeze probabilitatea ca cel puțin 2 aparate să nu facă apel sau probabilitatea ca numărul aparatelor care nu fac apel să fie cuprins între 1 și 3, inclusiv aceste numere.

$$R: \frac{1}{C_{34}^{6}} \sum_{k=2}^{6} C_{6}^{k} \cdot C_{28}^{6-k}$$

4. Se aruncă două monede. Să se calculeze probabilitatea să apară stema și banul?

$$R: \frac{1}{2}$$

5. În bibliotecă se așază 4 volume ale aceluiași autor. Care este probabilitatea ca acestea să fie aranjate în ordinea corectă?

$$R: \frac{1}{4!}$$

6. O urnă conține 5 bile albe și 3 negre. Se extrag 5 bile și se cere probabilitatea ca 3 să fie albe și două negre.

$$R: \frac{1}{C_{8}^{5}} \cdot C_{5}^{3} \cdot C_{3}^{2}$$

7. Se aruncă două zaruri. Se cere probabilitatea ca:
1° Suma punctelor să fie egală cu 9
2° „ „ „ cel puțin egală cu 9
3° „ „ „ cuprinsă între 6 și 10.

$$R: \frac{1}{9}, \frac{5}{18}, \frac{15}{36}$$

8. O urnă conține 3 bile albe, 4 negre și 5 roșii, iar alta conține 2 bile albe, 3 negre și 4 roșii. Se extrage din fiecare urnă cîte o bilă. Și se calculeze probabilitatea ca ambele bile să fie de aceeași culoare.

$$R: \frac{38}{108}$$

9. Într-o urnă sînt 5 bile albe și 7 negre. Se extrag 6 bile în bloc și apoi încă o bilă. Să se calculeze probabilitatea ca aceasta din urmă să fie neagră.

$$R: \frac{1}{6\, C_{12}^{6}} \sum_{k=0}^{5} C_{5}^{k}\, C_{7}^{6-k}\, (k+1)$$

Scanned with OKEN Scanner

10. Într-o urnă sînt 8 bile numerotate de la 1 la 8. Se extrag 5 bile și de exemplu a este cel mai mic număr. Să se calculeze probabilitatea P $(a \geqslant 2)$.

11. Se aruncă două zaruri. Să se calculeze probabilitatea ca suma punctelor ieșite să fie divizibilă prin 3, știind că primul zar a dat un număr cu soț.

$$R : \frac{1}{3}$$

12. În problema precedentă să se calculeze probabilitatea ca suma punctelor să fie 6, știind că suma acestor fețe a dat un număr cu soț.

R : 5/18

13. O urnă conține 5 bile albe și 4 negre. Să se calculeze probabilitatea ca după 4 extracții după schema bilei revenite să obținem 2 bile albe, știind că la extracția a doua a ieșit o bilă albă.

$$R : 3\left(\frac{5}{9}\right) - \left(\frac{4}{9}\right)^2$$

14. Să se rezolve problema precedentă în cazul cînd extracțiile se fac după schema bilei nerevenite.

R : 5/63

15. Se aruncă 12 zaruri. Să se determine probabilitatea ca fiecare față să apară cel puțin o dată.

16. Să se demonstreze formula (5′) .

17. Să se demonstreze inegalitatea lui Boole :

$$P\left(\bigcap_{k=1}^{s} A_k\right) \geqslant \sum_{k=1}^{s} P(A_k) - (s-1)$$

18) O piesă este standard dacă îndeplinește condițiile A, B, C. Din 10 000 piese 92 % îndeplinesc condiția A, 85 % condiția B și 89 % condiția C. Să se calculeze probabilitatea minimă ca o piesă să fie standard.

(R. După inegalitatea lui Boole $P \geqslant 0{,}66$

Capitolul IV

DISTRIBUȚII DISCRETE

§ 1. Variabile aleatorii

Să presupunem că avem cîmpul finit de probabilitate $[E,K,P]$ și $\{E_1, E_2, \ldots, E_n\}$ mulțimea evenimentelor elementare ale acestui cîmp. Să notăm în general prin e un eveniment oarecare din mulțimea $\{E_1, E_2, \ldots E_n\}$.

D e f i n i ț i e. Funcția reală

$$x = x\,(e) \tag{1}$$

determinată pentru orice eveniment elementar se numește *variabilă aleatorie definită pe mulțimea* $\{E_k\}$, $1 \leqslant k \leqslant n$.

Explicație. Să luăm pentru x_k o valoare fixă a funcției $x\,(e)$. Acestei valori îi corespunde în general o submulțime de evenimente elementare a căror reuniune formează un eveniment A_k al cîmpului [1]. Rezultă în mod evident $k \leqslant n$; de altă parte valorile x_k fiind distincte, evenimentele A_k sînt incompatibile două cîte două. În consecință evenimentele A_k pe care le presupunem în număr de $s \leqslant n$ constituie o desfacere a evenimentului sigur, deci

$$\bigcup_{k=1}^{s} A_k = E, \qquad A_k \cap A_h = \Phi \tag{1$_1$}$$

pentru $k, h = 1, 2, \ldots, s$; $k \neq h$.

Probabilitatea de a se realiza A_k este același lucru cu probabilitatea ca x să ia valoarea x_k, deci vom scrie

$$P\,(A_k) = P\,(x = x_k) \tag{1$_2$}$$

[1] Aceasta rezultă din inversarea funcției $x = x\,(e)$ din care deducem $e = e^{-1}\,(x)$

Această probabilitate depinzînd de x_k se mai notează astfel :

$$P\ (x = x_k) = f\ (x_k) \qquad (2)$$

Din relația (1_1) deducem :

$$P\left(\bigcup_{k=1}^{s} A_k\right) = \sum_{k=1}^{s} P\ (A_k) = \sum_{k=1}^{s} P\ (x = x_k) = 1$$

de unde :

$$\sum_{k=1}^{n} f\ (x_k) = 1 \qquad (2)$$

Exemple: 1°. Într-un lot de 18 becuri identice ca formă, 12 sînt albe și 6 albastre. Dacă evenimentul sigur este extracția unui bec oarecare din cele 18, vom avea mulțimea $\{E_1, E_2, \ldots . E_{18}\}$ de evenimente elementare. Să presupunem că becurile albe au fost numerotate 1, 2, ..., 12, iar cele albastre 13, 14, ..., 18.

Atunci funcția

$$x\ (e) = x\ (E_k) = \begin{cases} 0, \text{ dacă } 1 \leqslant k \leqslant 12 \\ 1, \text{ dacă } 13 \leqslant k \leqslant 18 \end{cases}$$

este o variabilă aleatorie în sensul definiției de mai sus. În adevăr funcția este determinată pentru orice eveniment elementar; de pildă $x\ (E_{10}) = 0$, $x\ (E_{16}) = 1$. La valoarea $x_k = 0$ corespund evenimentele $E_1, E_2, \ldots, E_{12}$ deci, un prim element al desfacerii $A_1 = \bigcup_{k=1}^{12} E_k$; la $x_k = 1$ corespund evenimentele $E_{13}, \ldots, E_{18}$, deci $A_2 = \bigcup_{k=13}^{18} E_k$. Evenimentul sigur admite o desfacere cu $s = 2$ evenimente incompatibile A_1 = extracția unui bec alb și A_2 = extracția unui bec albastru. Vom avea probabilitățile :

$$P\ (A_1) = P\ (x = 0) = f\ (0) = \frac{12}{18}$$

$$P\ (A_2) = P\ (x = 1) = f\ (1) = \frac{6}{18}$$

de unde :

$$f\ (0) + f\ (1) = 1 \qquad \text{(condiția } 2_1\text{)}$$

2°. Dacă se aruncă un zar avem mulțimea

$$\{E_k\}, \quad 1 \leqslant k \leqslant 6$$

unde se notează prin E_k apariția feței cu k puncte. Funcția :

$$x(E_k) = k, \qquad k = 1, 2, 3, 4, 5, 6$$

este o variabilă aleatorie. Desfacerea evenimentului sigur este formată din evenimentele elementare. $P(x = k) = f(k) = \frac{1}{6}$ pentru orice $k = 1, 2, \ldots 6$.

Se verifică imediat relația $\sum_{k=1}^{6} f(x) = 1$.

În mod analog, funcția

$$x(E_k) = \begin{cases} 0, \text{ pentru k impar} \\ 1, \text{ pentru k par} \end{cases}$$

este o variabilă aleatorie.

Valorii $x = 0$ îi corespund evenimentele E_1, E_3, E_5, iar valorii $x = 1$ evenimentele E_2, E_4, E_6. Avem deci desfacerea :

$A_1 = E_1 \cup E_3 \cup E_5$ = număr fără soț de puncte

$A_2 = E_2 \cup E_4 \cup E_6$ = număr cu soț de puncte

$$P(A_1) = P(x = 0) = P(E_1) + P(E_3) + P(E_5) = \frac{3}{6}$$

$$P(A_2) = P(x = 1) = P(E_2) + P(E_4) + P(E_6) = \frac{3}{6}$$

$$f(0) + f(1) = 1$$

3°. Se aruncă o monedă de un număr nelimitat de ori. Evenimentul sigur constă în apariția unui număr oarecare de steme. Atunci mulțimea evenimentelor elementare este $\{E_1, E_2, \ldots, E_n, \ldots\}$ unde E_1 este apariția unei steme într-o aruncătură, E_2 apariția stemei de două ori în două aruncături, ș.a.m.d.

Funcția :

$$x(E_k) = k, \qquad k = 1, 2, \ldots$$

este o variabilă aleatorie. Avem desfacerea $E = \bigcup_{k=1}^{\infty} E_k$

Probabilitatea ca din k probe să apară stema de k ori (eveniment a cărui probabilitate în fiecare probă este $p = \frac{1}{2}$) se calculează prin formula (3_7), capitolul III, în care facem $x = n = k$, $p = q = \frac{1}{2}$, deci :

$$P(E_k) = P(x = k) = f(k) = \frac{1}{2^k}, \quad k = 1, 2, \ldots$$

Relaţia (2_1) se verifică imediat. În adevăr, avem :

$$\sum_{k=1}^{\infty} f(k) = \sum_{k=1}^{\infty} \frac{1}{2^k} = \frac{\frac{1}{2}}{1 - \frac{1}{2}} = 1$$

Aşadar, valorile pe care le poate lua o variabilă aleatorie depind de întîmplare. Să presupunem că variabila aleatorie x ia valorile

$$0, 1, 2, \ldots, k, \ldots \tag{3}$$

cu probabilităţile respective

$$f(0), f(1), \ldots, f(k) \tag{3_1}$$

unde :

$$f(k) = P(x = k), \qquad k = 0, 1, 2, \ldots \tag{3_2}$$

Avem totdeauna îndeplinite condiţiile

$$f(k) \geqslant 0, \quad \sum f(k) = 1 \tag{3_4}$$

Mulţimea probabilităţilor $P(x = k) = f(k)$ se numeşte *distribuţia variabilei aleatorii* x sau simplu *distribuţie*, aşa încît de aici înainte vom folosi aproape exclusiv termenul de distribuţie, înţelegînd prin aceasta că s-a dat mulţimea probabilităţilor ca variabila să ia oricare din valorile sale posibile.

Aşa de pildă distribuţia din exemplul 1° de mai sus este

$$f(x) = \begin{cases} \frac{2}{3} & \text{dacă } x = 0 \\ \frac{1}{3}, & \text{dacă } x = 1 \end{cases}$$

Scanned with OKEN Scanner

În exemplul 2° avem respectiv distribuțiile

$$f(x) = \frac{1}{6}, \qquad x = 1, 2, \ldots, 6$$

$$f(x) = \frac{1}{2}, \qquad \text{dacă } x = 0, 1$$

În sfîrșit, pentru distribuția din exemplul 3° avem :

$$f(x) = \frac{1}{2^x}, \qquad x = 1, 2, \ldots$$

§ 2. Distribuții unidimensionale

Prin definiție, dacă variabila x ia un număr finit sau o infinitate numărabilă de valori, distribuția se numește distribuție de *tip discret* sau, mai simplu, *distribuție discretă*. Probabilitatea ca variabila x să ia o anumită valoare x am notat-o prin f (x), adică :

$$P\{x(e) = x\} = f(x) \tag{4}$$

Funcția f (x) o vom numi *densitatea de probabilitate* a distribuției sau *funcția de densitate*[1]. Ea verifică relațiile cunoscute (3_4) :

$$f(x) \geqslant 0, \quad \sum f(x) = 1 \tag{4_1}$$

unde însumarea se face asupra tuturor valorilor pe care le poate lua variabila x (e).

Cunoașterea densității de probabilitate f (x) joacă un rol fundamental în studiul distribuțiilor statistice. Vom arăta aceasta pe un exemplu simplu. Să considerăm un lot de 2 becuri, unul de 60 lumini și altul de 75 lumini, identice ca formă și culoare și să efectuăm 4 extracții după schema bilei revenite. Fie x variabila aleatorie care indică numărul de cîte ori a ieșit becul de 60 lumini în cele 4 extracții. Avem distribuția :

$$f(x) = \frac{C_4^x}{2^4} \qquad x = 0, 1, 2, 3, 4$$

[1] Noțiunea de densitate este legată de distribuțiile continue ; în scopul uniformizării terminologiei am folosit această denumire și pentru distribuțiile discrete.

Condițiile (4_1) se verifică imediat : condiția $f(x) > 0$ este evidentă, iar $\sum_{x=0}^{4} f(x) = \frac{1}{2^4} \sum_{x=0}^{4} C_4^x = \frac{1}{2^4} \cdot 2^4 = 1$

deoarece :

$$(1 + 1)^4 = \sum_{x=0}^{4} C_4^x = 2^4$$

Să rezolvăm cîteva probleme în legătură cu această distribuție :

1°. Probabilitatea ca numărul de cîte ori a ieșit becul de 60 de lumini să fie mai mic decît 3 este, în baza regulei de adunare a probabilităților :

$$P(x < 3) = f(0) + f(1) + f(2) = \frac{1}{2^4} + \frac{4}{2^4} + \frac{6}{2^4} = \frac{11}{16}$$

2°. Probabilitatea ca acest număr să fie cuprins între 1 și 3 inclusiv va fi :

$$P(1 \leqslant x \leqslant 3) = f(1) + f(2) + f(3) = \frac{4}{2^4} + \frac{6}{2^4} + \frac{4}{2^4} = \frac{7}{8}$$

3°. Știind că becul de 60 lumini a apărut mai puțin de 4 ori, probabilitatea ca el să nu iasă cel mult de două ori este probabilitatea condiționată :

$$P_{x<4}(x \leqslant 2) = P(x \leqslant 2 / x < 4) = \frac{P(x \leqslant 2)}{P(x < 4)}$$

Avem însă :

$$P(x \leqslant 2) = f(0) + f(1) + f(2) = \frac{1}{2^4} + \frac{4}{2^4} + \frac{6}{2^4} = \frac{11}{16}$$

$$P(x < 4) = f(0) + f(1) + f(2) + f(3) =$$

$$= \frac{1}{2^4} + \frac{4}{2^4} + \frac{6}{2^4} + \frac{4}{2^4} = \frac{15}{16}$$

de unde :

$$P_{x<4}(x \leqslant 2) = \frac{\sum_{x=0}^{2} f(x)}{\sum_{x=0}^{3} f(x)} = \frac{\frac{11}{16}}{\frac{15}{16}} = \frac{11}{15}$$

Scanned with OKEN Scanner

Funcția de repartiție a unei distribuțiii este probabilitatea ca variabila x (e) să ia o valoare mai mică decît o anumită valoare dată x. Această probabilitate este evident o funcție de valoarea x a variabilei ; o vom nota prin F (x).

În baza definiției avem relația

$$F(x) = P[x(e) < x] \tag{5}$$

Dacă variabila x (e) ia valorile (3) avem evident :

$$F(k) = P(x < k) = f(0) + f(1) + \ldots\ldots + f(k-1) = \sum_{x=0}^{k-1} f(x)$$

Mai general, dacă x ia valorile $x_0 < x_1 < x_2 < \ldots < x_s$ funcția de repartiție este dată de relația :

$$F(x_k) = P(x < x_k) = \sum_{x < x_k} f(x) \tag{5_1}$$

unde suma se extinde asupra tuturor valorilor $x_0, x_1, \ldots, x_{k-1}$ mai mici decît x_k .

Observație. Din (5_1) deducem $F(x_{k+1}) = \sum_{x < x_{k+1}} f(x)$ de unde prin scădere se găsește :

$$f(x_k) = F(x_{k+1}) - F(x_k) \tag{5_2}$$

Așadar, cunoscînd funcția de repartiție F (x) se poate determina densitatea f (x) a distribuției așa încît o distribuție se consideră dată fie prin densitatea sa, fie prin funcția de repartiție. Să calculăm de exemplu funcția de repartiție a distribuției

$$f(x) = \frac{C_4^x}{2^4} \qquad x = 0, 1, 2, 3, 4$$

Vom avea :

$$F(0) = P(x < 0) = 0$$

$$F(1) = P(x < 1) = f(0) = \frac{1}{16}$$

$$F(2) = P(x < 2) = f(0) + f(1) = 5/16$$

$$F(3) = P(x < 3) = f(0) + f(1) + f(2) = 11/16$$

$$F(4) = P(x < 4) = f(0) + f(1) + f(2) + f(3) = \frac{15}{16}$$

$$F(x) = \sum_{x=0}^{4} f(x) = 1 \text{ pentru } x \geqslant 4.$$

Scanned with OKEN Scanner

Aceste valori se scriu de obicei sub formă de tablou :

$$F(x) = \begin{cases} 0 & \text{pentru } x < 0 \\ 1/16 & ,, \quad 0 \leqslant x < 1 \\ 5/16 & ,, \quad 1 \leqslant x < 2 \\ 11/16 & ,, \quad 2 \leqslant x < 3 \\ 15/16 & ,, \quad 3 \leqslant x < 4 \\ 1 & ,, \quad 4 \leqslant x \end{cases}$$

Din cele arătate pînă acum rezultă că probabilitatea ca variabila x să ia o valoare oarecare cuprinsă în intervalul (k, h) se obține însumînd densitatea de probabilitate pentru toate valorile variabilei din acest interval, adică :

$$P(k \leqslant x < h) = \sum_{x=k}^{h-1} f(x) = f(k) + f(k+1) + \ldots + f(h-1) \qquad (5_3)$$

sau încă ținînd seama de (5_2)

$$P(k \leqslant x < h) = F(h) - F(k) \qquad (5_1)$$

Adică probabilitatea ca x să ia o valoare din intervalul (k, h) se obține făcînd diferența valorilor funcției de repartiție la extremitățile intervalului.

Vom descrie acum cîteva din distribuțiile de tip discret care sînt mai des întîlnite în problemele practice.

1°. *Distribuția uniformă* este caracterizată prin densitatea de probabilitate

$$f(x) = \frac{1}{n}, \qquad x = 1, 2, \ldots, n \qquad (6)$$

O asemenea distribuție se obține de pildă cînd se aruncă un zar, $(n = 6)$, $f(x) = \frac{1}{6}$, deci : $\sum_{x=1}^{6} f(x) = 1$

2°. *Distribuția binomială.* Să considerăm o populație compusă din N elemente. Dintre acestea a posedă o anumită caracteristică A, iar restul nu o posedă ; se extrag n elemente întorcînd de fiecare dată elementele cercetate în colectivitate.

O atare distribuție este caracterizată prin densitatea

$$f(x) = C_n^x \, p^x \, q^{n-x} = \frac{n!}{x!\,(n-x)!} \, p^x \, q^{n-x} \qquad (7)$$

$$0 \leqslant x \leqslant n$$

unde x reprezintă numărul elementelor cu caracteristica A din cele n extrase, iar $p = P(A) = \frac{a}{N}$, $q = 1 - p$. Densitatea f(x) s-a calculat, evident, după schema bilei revenite [vezi relația (3_7), capitolul III]. Condițiile (4_1) se verifică ușor :

$$f(x) \geqslant 0 ; \sum_{x=0}^{n} f(x) = \sum_{x=0}^{n} C_n^x p^x q^{n-x} = (p+q)^n = 1$$

Densitatea f (x) depinde de numerele n, și p numite *parametrii distribuției* (q nu este un parametru esențial deoarece $q = 1 - p$). Se spune că distribuția formează o *familie dublu parametrică*. Parametrul n este un parametru discret, el poate lua numai valori întregi și pozitive, în timp ce p este un parametru continuu care poate lua orice valoare din intervalul închis [1,0].

Funcția f (x) definită numai pentru valori întregi și pozitive ale lui x pe segmentul [0, n], crește pînă la valoarea $x = m$ și apoi descrește pînă la valoarea $x = n$ a variabilei. Valoarea $x = m$ se numește *valoarea modală* sau *valoarea cea mai probabilă* a ditribuției.

Este evident că f (x) va crește atîta timp cît raportul $p(x) = f(x+1)/f(x)$ este mai mare ca 1. Avînd relația :

$$\frac{f(x+1)}{f(x)} = \frac{(n-x)\,p}{(x+1)\,q} \qquad (7_1)$$

obținem $x = np + p = (n+1)\,p$. Dacă $x > (n+1)\,p$, funcția f(x) descrește. În general, numărul $(n+1)\,p$ este o fracție zecimală, pentru care vom lua $m = [(n+1)\,p]$, adică partea întreagă a numărului $(n+1)\,p$[1]. Dacă $(n+1)\,p$ este întreg, f(x) are două maxime consecutive egale.

În general putem scrie inegalitățile $(n+1)\,p - 1 \leqslant m$ $m \leqslant (n+1)\,p$ sau încă înlocuind $1 - p = q$

$$np - q \leqslant m \leqslant (n+1)\,p \qquad (7_2)$$

D e f i n i ț i e : Mulțimea de n elemente extrase din colectivitatea inițială de N elemente, se numește *selecție repetată de mărimea n.*

Densitatea f (x) se poate reprezenta geometric în planul a două axe perpendiculare pentru diferite valori ale lui n și p în

[1] Funcția [x] se numește partea întreagă a lui x, iar x — [x] partea fracționară

modul următor : valorile lui x se așază pe axa absciselor, iar valorile densității f (x) se reprezintă prin segmente paralele cu axa oy.

În graficul de mai jos am reprezentat funcția f (x) pentru $n = 4$, $p = \frac{1}{2}$.

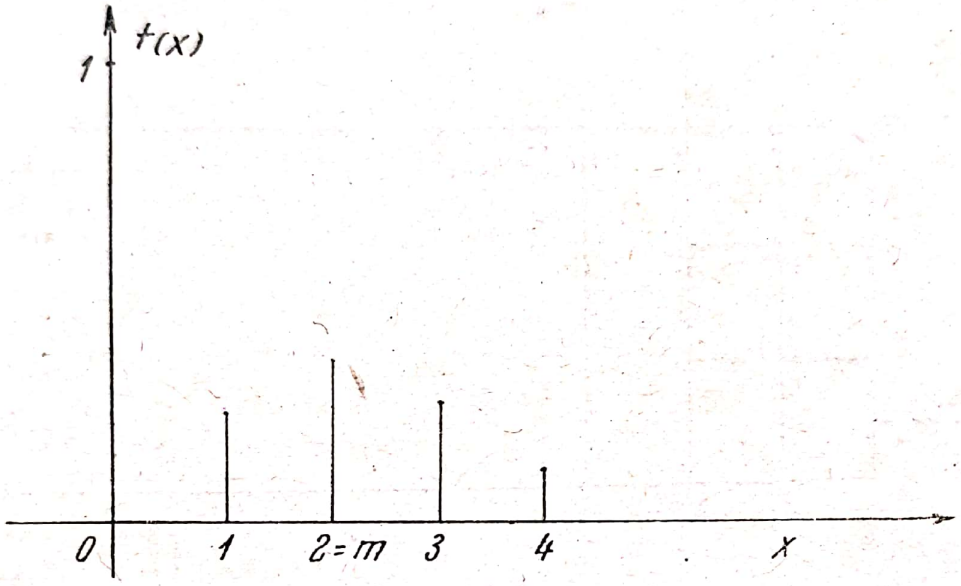

Fig. 17

Punctul M (x) de pe axa 0 x a cărui abscisă ia valorile x = 0, 1 2, 3, 4 pe care le ia și variabila aleatorie x se numește *punct aleator*. Uneori este mai convenabil să folosim noțiunea de punct aleator în locul variabilei aleatorii. Așa de pildă, în loc de probabilitatea ca variabila x să ia o valoare cuprinsă între 1 și 3 inclusiv, vom spune probabilitatea ca punctul aleator să se găsească pe segmentul [1, 3]. După cum am văzut mai înainte avem : $P(M \in [1, 3)[= P(1 \leqslant x \leqslant 3) = 7/8$
Știind că punctul se găsește în intervalul [0, 4], probabilitatea ca el să nu fie la dreapta punctului de abscisă 2 (adică să se găsească pe segmentul [0, 2] este : $P(M \in [0, 2]/M \in [0,4]) = P(x \leqslant 2/x < 4) = 11/15$
Funcția de repartiție a distribuției binomiale este :

$$F(k) = P(x = k) = \sum_{x=0}^{k} C_n^x \, p^x \, q^{n-x}$$

Mai înainte am calculat valorile funcției de repartiție F (x)

pentru $x = 4$, $p = \frac{1}{2}$. Grafic, funcția de repartiție pentru acest caz particular se reprezintă ca în figura alăturată :

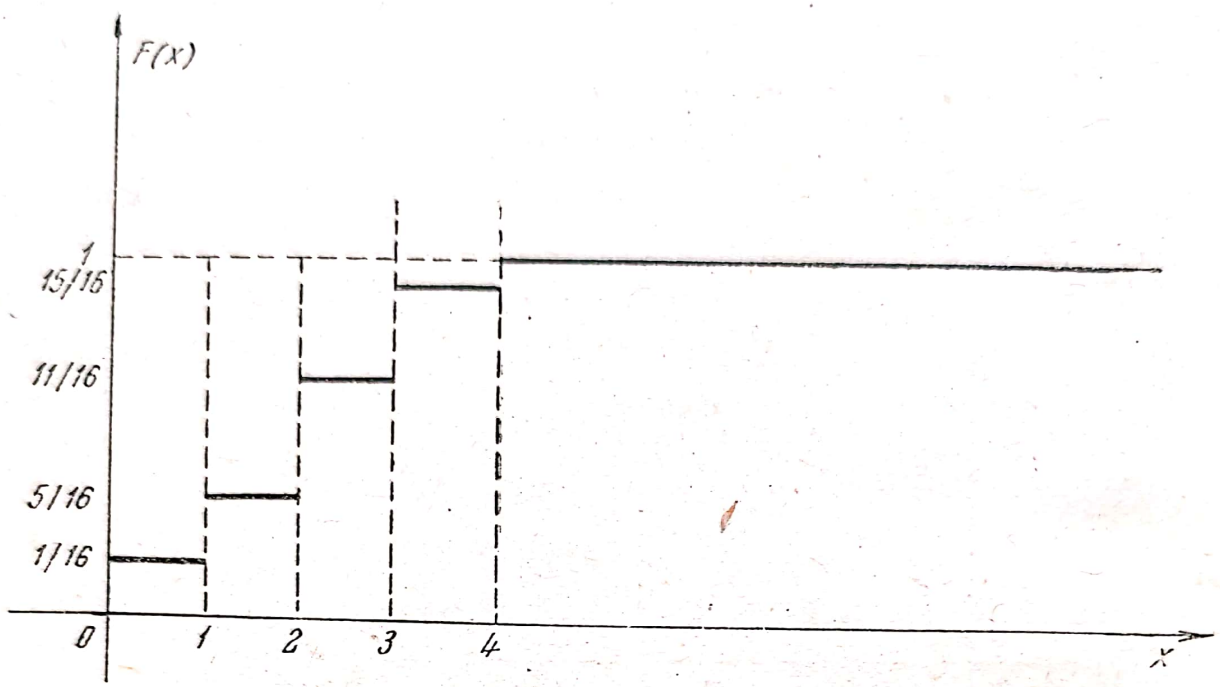

Fig. 18

În interpretarea geometrică dată mai sus, funcția de repartiție reprezintă probabilitatea ca punctul aleator M de pe axa 0x să aparțină intervalului [a, x), unde a este cea mai mică valoare a variabilei aleatorie, deci :

$$F(x) = P(M \in [a, x])$$

Observații : 1°. Densitățile f (x) ale distribuției binomiale sînt egale cu termenii dezvoltării binomului $(p + q)^n$. De aici și numele distribuției.

2°. Calculul densității f (x) pentru valori mari ale lui n este foarte anevoios din cauza factorialelor cu care exprimăm combinările ; de aceea în aplicații se folosește adesea pentru calculul acestor factoriale formula aproximativă a lui Stirling :

$$n! = n^n \, e^{-n} \sqrt{2\pi n} \qquad (7_4)$$

unde π și e au semnificațiile cunoscute.

3°. *Distribuția lui Poisson* are densitatea

$$f(x) = e^{-a} \frac{a^x}{x!}, \qquad x = 0, 1, 2, \ldots \qquad (8)$$

Scanned with OKEN Scanner

unde $a > 0$. Se vede ușor că prima condiție (4_1) este îndeplinită. Pentru verificarea condiției a doua vom scrie :

$$\sum_{x=0}^{\infty} f(x) = e^{-a} \sum_{x=0}^{\infty} \frac{a^x}{x!} = e^{-a} \cdot e^{a} = 1$$

Această distribuție se obține ca o distribuție limită a celei binomiale. În adevăr să notăm în distribuția $f(x) = C_n^x p^x q^{n-x}$:

$$np = a \qquad (8_1)$$

unde a este un număr finit și să calculăm densitatea $f(x)$ pentru $n \to \infty$ și x finit [1]. Vom putea scrie

$$f(x) = \frac{n(n-1)\ldots(n-x+1)}{x!} p^x q^{n-x} =$$

$$= \frac{n(n-1)\ldots(n-x+1)}{n^x} \cdot \frac{n^x p^x}{x!} q^{n-x} =$$

$$= \frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{(n-x+1)}{n} \cdot \frac{a^x}{x!} \left(1 - \frac{a}{n}\right)^{n-x} =$$

$$= \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{x-1}{n}\right) \frac{a^x}{x!} \frac{\left(1 - \frac{a}{n}\right)^n}{\left(1 - \frac{a}{n}\right)^x}$$

Pentru $n \to \infty$ avem limitele :

$$\lim_{n \to \infty} \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{x-1}{n}\right) = 1$$

$$\lim_{n \to \infty} \left(1 - \frac{a}{n}\right)^n = e^{-a} ; \quad \lim_{n \to \infty} \left(1 - \frac{a}{n}\right)^x = 1$$

Obținem deci distribuția limită $f(x) = e^{-a} \frac{a^x}{x!}$, adică distribuția lui Poisson. Această distribuție are importante aplicații.

4°. *Distribuția hipergeometrică* este caracterizată prin densitatea de probabilitate

$$f(x) = \frac{C_a^x C_{N-a}^{n-x}}{C_N^n} \qquad (9)$$

[1] Se spune că evaluăm asimptotic pe $f(x)$.

Avem evident $f(x) \geqslant 0$. Să arătăm că suma $\sum_{x=0}^{n} f(x) = 1$.

Pentru aceasta vom scrie :

$$\sum_{x=0}^{n} f(x) = \frac{1}{C_N^n} \sum_{x=0}^{n} C_a^x \cdot C_{N-a}^{n-x}$$

și va trebui să demonstrăm relația următoare :

$$\sum_{x=0}^{n} C_a^x \cdot C_{N-a}^{n-x} = C_N^n \qquad (9_1)$$

Pornim de la identitatea : $(1+x)^{a+b} = (1+x)^a \cdot (1+x)^b$ sau încă dezvoltînd fiecare binom :

$$1 + C_{a+b}\, x + C_{a+b}^2\, x^2 + \ldots + x^{a+b} =$$
$$= \left(1 + C_a^1\, x + C_a^2\, x + \ldots + x^a\right)\left(1 + C_b^1\, x + \ldots + x^b\right)$$

Prin identificare deducem :

$$C_{a+b}^1 = C_a^1 + C_b^1 = \sum_{x=0}^{1} C_a^x \cdot C_b^{1-x}$$

$$C_{a+b}^2 = C_a^1 \cdot C_b^1 + C_a^2 + C_b^2 = \sum_{x=0}^{2} C_a^x\, C_b^{2-x}$$

. . . . . .

$$C_{a+b}^n = \sum_{x=0}^{n} C_a^x\, C_b^{n-x}$$

Făcînd în ultima relație $b = N - a$ obținem (9_1). Distribuția hipergeometrică se întîlnește în probleme de genul următor :

Dintr-o populație compusă din N elemente din care a posedă caracteristica A și restul N—a nu o posedă, se extrag deodată N elemente (sau ceea ce este tot una se extrag succesiv n elemente fără a reintroduce în colectivitate elementul cercetat). Să determinăm probabilitatea ca x din cele n elemente extrase să posede caracteristica A și (n—x) să nu o posede. O asemenea probabilitate se calculează cu formula (2_2), capitolul III, adică după schema bilei nerevenite.

D e f i n i ț i e. Mulțimea de n elemente extrasă din colectivitatea inițială în modul arătat mai sus se numește *selecție nerepetată de mărimea n.*

Scanned with OKEN Scanner

§ 3. Distribuții multidimensionale

Fie x și y două variabile aleatorii avînd respectiv densitățile de probabilitate f (x) și φ (y). Variabila x determină desfacerea $A_1, A_2, \ldots, A_r$ a evenimentului sigur în elemente incompatibile, iar variabila y determină desfacerea $B_1, B_2, \ldots, B_s$. După definiția desfacerii avem relațiile :

$$\bigcup_{k=1}^{r} A_k = E,\ A_k \cap A_h = \Phi;\ k, h = 1, 2, \ldots, r;\ k \neq h$$

$$\bigcup_{k=1}^{s} B_k = E,\ B_k \cap B_h = \Phi;\ k, h = 1, 2, \ldots, s;\ k \neq h$$

Pentru valoarea x_k a lui x corespunde evenimentul A_k $(1 \leqslant k \leqslant r)$, iar pentru $y = y_h$ corespunde evenimentul B_h $(1 \leqslant h \leqslant s)$. Vom considera ca evenimente unice, evenimentele mixte $A_k \cap B_h = C_{k_h}$ $(k = 1, 2, \ldots, r;\ h = 1, 2, \ldots, s)$ în număr de r s. Evenimentele $C_{k\,h}$ sînt evident incompatibile două cîte două.

Probabilitatea evenimentului $C_{k\,h}$:

$$P\ (C_{kh}) = P\ (A_k \cap B_h) \tag{10}$$

este tot una cu probabilitatea ca variabila x să ia valoarea x_k și variabila y să ia valoarea y_h. Notînd această probabilitate prin $f\ (x_k, y_h)$ avem :

$$P\ (A_k \cap B_h) = P\ (x = x_k,\ y = y_h) = f\ (x_k, y_h) \tag{10_1}$$

Funcția $f\ (x_k, y_h)$ este evident nenegativă. Deoarece evenimentele $A_k \cap B_h$ sînt incompatibile suma :

$$\sum_{k,\,h} P\ (C_{k\,h}) = \sum_{k=1}^{r} \sum_{h=1}^{s} f\ (x_k, y_h) = 1$$

Suma dublă $\sum_{k=1}^{r} \sum_{h=1}^{s} f\ (x_k, y_h)$ se dezvoltă în etape. Vom putea scrie succesiv :

$$\sum_{k=1}^{r} \sum_{h=1}^{s} f\ (x_k, y_h) = \sum_{k=1}^{r} \{ f\ (x_k, y_1) + f\ (x_k, y_2) + \ldots + f\ (x_k, y_s) \} =$$

$$= \sum_{k=1}^{r} f\ (x_k, y_1) + \sum_{k=1}^{r} f\ (x_k, y_2) + \ldots + \sum_{=}^{r} f\ (x_k, y_s) =$$

$$= f(x_1, y_1) + f(x_2, y_1) + \ldots + f(x_r, y_1) + \quad (10_2)$$
$$+ f(x_1, y_2) + f(x_2, y_2) + \ldots + f(x_r, y_2) +$$
$$+ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$+ f(x_1, y_s) + f(x_2, y_s) + \ldots + f(x_r, y_s)$$

Funcția $f(x_k, y_h)$ de două variabile se numește densitate de probabilitate asociată celor două variabile aleatorii x și y.

Așadar, unui sistem de două variabile aleatorii x (e) și y (e) îi corespunde o distribuție caracterizată prin densitatea de probabilitate

$$f(x, y) = P[x(e) = x, \; y(e) = y] \quad (10_3)$$

cu proprietățile

$$f(x, y) \geqslant 0, \; \sum_x \sum_y f(x, y) = 1 \quad (10_4)$$

unde suma dublă se extinde asupra tuturor valorilor posibile ale variabilelor x și y.

O asemenea distribuție se numește *bidimensională.*

Exemple : 1°. Să considerăm distribuția unui număr de 57 becuri repartizat după numărul de lumini și după culoare :

Tabloul 15

Nr. de lumini	Albe	Albastre	Total
65	23	9	32
70	14	5	19
120	4	2	6
Total	41	16	57

Variabilele x și y sînt numărul de lumini, respectiv culoarea becului. Să notăm A_1 evenimentul ca un bec să aibă 60 de lumini, A_2 să aibă 75 și A_3 să aibă 120 lumini ; în mod analog B_1 = evenimentul ca becul să fie alb și B_2 să fie albastru. În acest caz $r = 3$, $s = 2$.

Să facem ca evenimentului A_1 să-i corespundă 0, lui A_2 să-i corespundă 1 și lui A_3 să-i corespundă numărul 2. Vom asocia de asemenea lui B_1 pe zero, iar lui B_2 pe 1.

Densitatea f (x, y) este definită pentru $x = 0, 1, 2$, $y = 0,1$ Avem de pildă valorile : $f(2, 1) = \frac{2}{57}$, $f(0,1) = \frac{9}{57}$.

Scanned with OKEN Scanner

Să verificăm condiția a doua din relația (1_4), adică :

$$\sum_{x=0}^{2}\sum_{y=0}^{1} f(x,y) = 1.$$

Dezvoltînd după relația (10_2) găsim :

$$\sum_{x=0}^{2}\sum_{y=0}^{1} f(x,y) = \sum_{x=0}^{2}[f(x,0) + f(x,1)] = \sum_{x=0}^{2} f(x,0) + \sum_{x=0}^{2} f(x,1) =$$

$$= f(0,0) + f(1,0) + f(2,0) + f(0,1) + f(1,1) + f(2,1) =$$

$$= \frac{23}{57} + \frac{14}{57} + \frac{4}{57} + \frac{9}{57} + \frac{5}{57} + \frac{2}{57} = \frac{57}{57} = 1$$

Geometric, distribuția bidimensională se reprezintă în modul următor : considerăm două axe perpendiculare ox și oy în planul orizontal ; pe ox figurăm valorile 0, 1, 2, ale lui x, iar pe oy valorile 0 și 1 ale lui y. În punctele M (x, y) din plan ridicăm perpendiculare pe plan, egale cu valorile densității respective f (x, y).

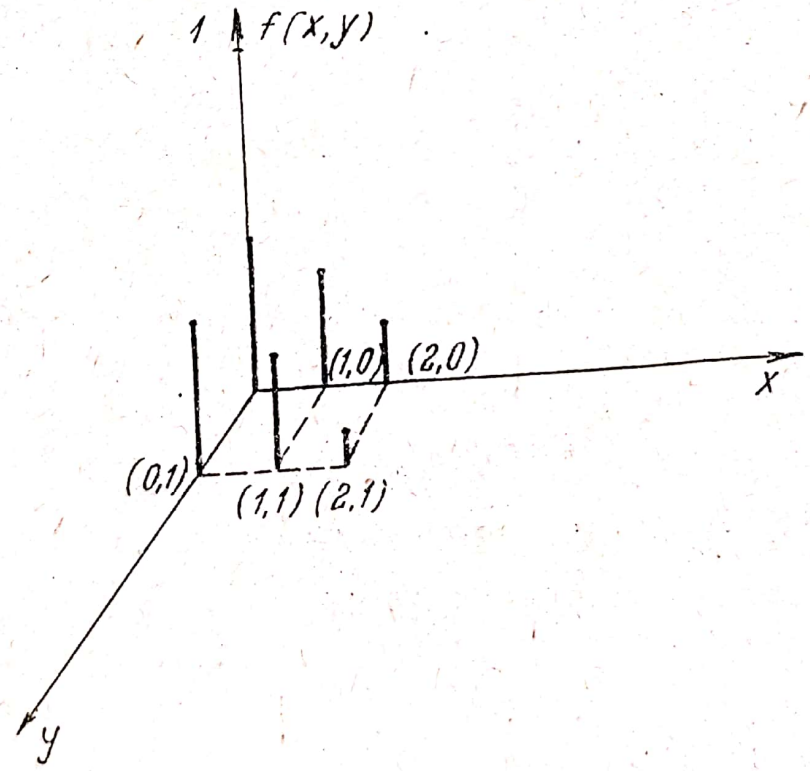

Fig. 19

Perpendiculara ridicată în punctul (0, 0) de exemplu este egală cu 23/57, ..., cea ridicată în punctul (2, 1) este 2/57.

2°. O urnă conține 5 bile albe, 3 negre și 6 roșii. Se extrag în bloc 7 bile și se cere probabilitatea ca x să fie albe și y să fie negre (evident $7 - x - y$ vor fi roșii). Probabilitatea căutată se calculează cu formula (2_3), capitolul III, adică :

$$f(x, y) = \frac{C_5^x \cdot C_3^y \cdot C_6^{7-x-y}}{C_{14}^7}$$

$$0 \leqslant x \leqslant 5\,;\ 0 \leqslant y \leqslant 3\,;\ x + y \leqslant 7$$

Am obținut din nou o distribuție bidimensională. Densitatea $f(x, y)$ este definită în punctele (x, y) din dreptunghiul $x = 0$, $x = 5$, $y = 0$, $y = 3$, considerînd $f(x, y) = 0$ pentru valorile $x > 5$, $y > 3$ sau $x + y \geqslant 7$.

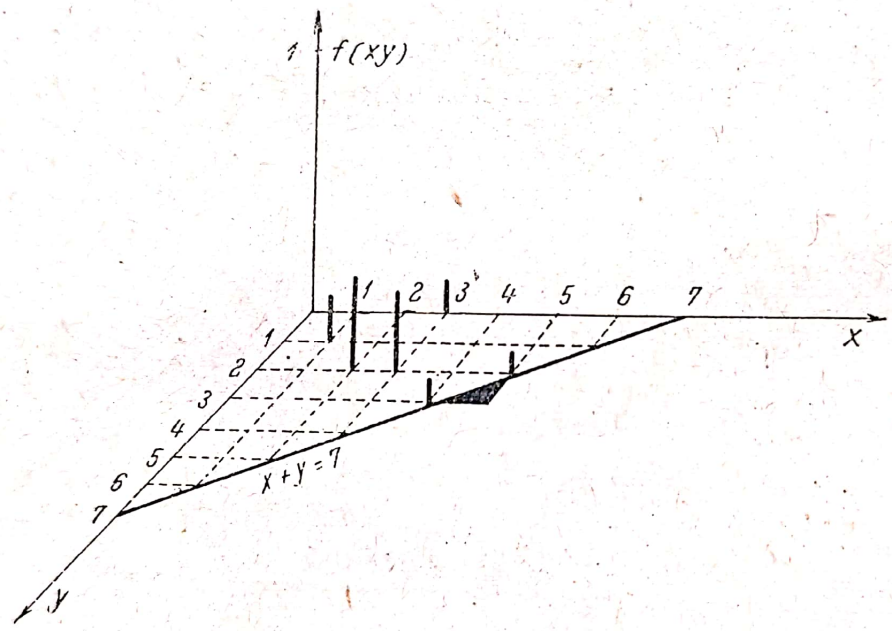

Fig. 20

Dăm mai jos cîteva valori ale funcției $f(x, y)$:

$$f(0,0) = \frac{C_5^0 \cdot C_3^0 \cdot C_6^7}{C_{14}^7} = 0 \quad \text{căci} \quad C_6^7 = 0$$

$$f(1,1) = \frac{C_5^1 \cdot C_3^1 \cdot C_6^5}{C_{14}^7} = \frac{90}{3\,432}$$

$$f(2,2) = \frac{C_5^2 \cdot C_3^2 \cdot C_6^3}{C_{14}^7} = \frac{600}{3\,432}$$

Scanned with OKEN Scanner

$$f(3,2) = \frac{C_5^3 \cdot C_3^2 \cdot C_6^2}{C_{14}^7} = \frac{450}{3\,432}$$

$$f(4,3) = \frac{C_5^4 \cdot C_3^3 \cdot C_6^0}{C_{14}^7} = \frac{5}{3\,432}$$

$$f(5,2) = \frac{C_5^5 \cdot C_3^2 \cdot C_6^0}{C_{14}^7} = \frac{3}{3\,432}$$

$$f(0,3) = \frac{C_5^0 \cdot C_3^3 \cdot C_6^4}{C_{14}^7} = \frac{15}{3\,432}$$

Toate perpendicularele în punctele (x, y) pe planul xy, reprezentînd valorile densității f (x, y) sînt situate în dreptunghiul de definiție al funcției între axa verticală și planul paralel cu ea, $x + y = 7$.

Observație. Dacă notăm prin z numărul bilelor roșii din cele 7 extrase obținem o distribuție aparent tridimensională echivalentă cu cea bidimensională, avînd densitatea de probabilitate

$$f(x, y, z) = \frac{C_5^x \cdot C_3^y \cdot C_6^z}{C_{14}^7}$$

$0 \leqslant x \leqslant 5$; $0 \leqslant y \leqslant 3$; $0 \leqslant z \leqslant 6$; $x + y + z = 7$

Geometric, o asemenea distribuție trebuie să o prezentăm în spațiul cu 4 dimensiuni și această reprezentare nu mai este intuitivă. În aceste cazuri ne folosim de expresia analitică a funcției de densitate. Ca și în cazul unidimensional, densitatea de probabilitate joacă un rol foarte important în rezolvarea multora din problemele ce se pun în legătură cu distribuțiile statistice.

Așa de pildă probabilitatea ca variabilele x și y să ia valori într-un anumit domeniu în care funcția f (x, y) este definită, se va obține însumînd f (x, y) pentru punctele acelui domeniu. Folosind limbajul geometric vom spune că probabilitatea ca punctul aleatoriu M (x, y) să se găsească într-un anumit domeniu se obține însumînd funcția de densitate în toate punctele domeniului în care M se poate găsi.

Așa de exemplu, probabilitatea ca punctul aleatoriu M (x, y) să apară în dreptunghiul $0 \leqslant x \leqslant 3$, $0 \leqslant y \leqslant 2$, va fi:

$$P(0 \leq x \leq 3, \quad 0 \leq y \leq 2) = \sum_{x=0}^{3} \sum_{y=0}^{2} f(x,y)$$

Pentru întregul domeniu de definiție al distribuției, avem după cum se știe :

$$\sum_{x=0}^{5} \sum_{y=0}^{3} f(x, y) = 1$$

În general să presupunem că $x = 0, 1, 2, \ldots, r$; $y = 0, 1, 2, \cdots s$, deci funcția $f(x, y)$ este definită pe dreptunghiul $x = 0$, $x = r$, $y = 0$, $y = s$. Probabilitatea ca $x \leqslant 4$, $y < 3$ va fi :

$$P(x \leqslant 4, y < 3) = \sum_{x=0}^{4} \sum_{y=0}^{2} f(x, y)$$

Probabilitatea $P(x + y < 4)$ se va obține însumînd $f(x, y)$ în interiorul triunghiului $x = 0$, $y = 0$, $x + y = 4$, deci :

$$P(x+y<4) = \sum_{x=0}^{3} \sum_{y=0}^{3-x} f(x, y) = \sum_{y=0}^{3} \sum_{x=0}^{3-y} f(x, y) =$$
$$= f(0,0) + f(1,0) + f(2,0) + f(3,0)$$
$$+ f(0,1) + f(1,1) + f(2,1)$$
$$+ f(0,2) + f(1,2)$$
$$+ f(0,3)$$

Pentru a calcula probabilitatea $P(x + y = 4)$ va trebui să însumăm $f(x, y)$ pentru toate punctele care se găsesc pe dreapta $x + y = 4$, deci :

$$P(x + y = 4) = \sum_{x=0}^{4} f(x, 4-x) = \sum_{y=0}^{4} f(4 - y, y) =$$
$$= f(0,4) + f(1,3) + f(2,2) + f(3,1) + f(4,0)$$

În mod analog obținem probabilitățile condiționate :

$$\underset{y=3}{P}(x \leq 2) = \sum_{x=0}^{2} \underset{y=3}{P}[x(e) = x] = \frac{\sum_{x=0}^{2} f(x, 3)}{\sum_{x=0}^{r} f(x, 3)}$$

$$\underset{y>3}{P}(x \leq 2) = \frac{\sum_{x=0}^{2} \sum_{y=4}^{s} f(x, y)}{\sum_{x=0}^{r} \sum_{y=4}^{s} f(x, y)}$$

Scanned with OKEN Scanner

Pentru o distribuție de trei variabile cu densitatea f (x, y, z) unde

$$f(x, y, z) \geqslant 0, \sum_x \sum_y \sum_z f(x, y, z) = 1$$

avem probabilitățile :

$$P(x \leq 3, y \leq 4, 2 \leq z \leq 6) = \sum_{x=0}^{3} \sum_{y=0}^{4} \sum_{z=2}^{6} f(x, y, z)$$

În mod analog, obținem altă probabilitate și anume :

$$P(x + y + z \leqslant 4). = \sum_{x=0}^{4} \sum_{y=0}^{4-x} \sum_{z=0}^{4-x-y} f(x, y, z)$$

Să considerăm acum s variabile aleatorii discrete

$$x_1, x_2, \ldots, x_s \tag{11}$$

cărora le asociem densitatea de probabilitate

$$P[x_1(e) = x_1, x_2(e) = x_2, \ldots x_s(e) = x_s] = f(x_1, x_2, \ldots\ldots, x_s) \tag{11_1}$$

cu proprietățile :

$$f(x_1, x_2, \ldots\ldots, x_s) \geq 0, \sum_{x_1=0}^{r_1} \sum_{x_2=0}^{r_2} \ldots \sum_{x_s=0}^{r_s} f(x_1, x_2, \ldots\ldots x_s) = 1 \tag{11_2}$$

unde am presupus că variabila x_k ia valorile :

$$0, 1, 2, \ldots, r_k ; 1 \leq k \leq s \tag{11_3}$$

Problemele ce se pun în legătură cu această distribuție *s — dimensională* se rezolvă în mod analog cu cele tratate în cazul particular s = 2 sau s = 3.

Distribuții marginale. Să considerăm din nou o distribuție bidimensională cu densitatea f (x, y). În unele probleme de statistică se poate să ne intereseze numai variabila x, în altele numai y, indiferent de ce valori ia cealaltă variabilă. Se numește *distribuție marginală* în raport cu x probabilitatea

$$P[x(e) < x] = f(x) \tag{12}$$

indiferent de valorile luate de y. Să reprezentăm grafic distribuția dată, față de un sistem de două axe perpendiculare ca în fig. 21.

Funcția f (x, y) = P [x (e) = x, y (e) = y] reprezintă probabilitatea ca punctul aleator M să aibă precis coordonatele

Scanned with OKEN Scanner

x și y. Pe verticala punctului x, variabila y poate lua valorile y = 0, 1, 2, ..., s; punctul aleatoriu M va avea succesiv coordonatele (x, 0), (x, 1), ... (x, s), deci întotdeauna abscisa x.

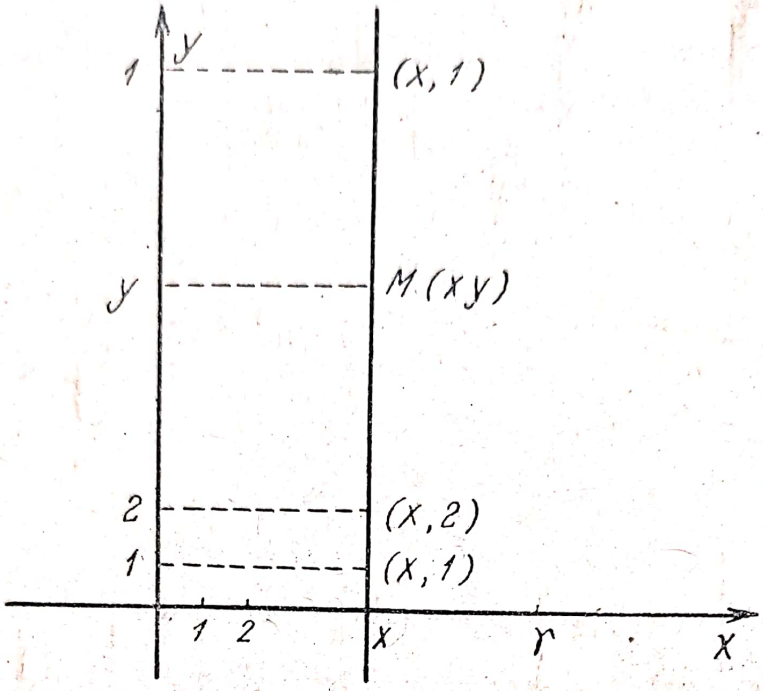

Fig. 21

Deoarece punctul M ocupă poziții diferite, evenimentele $\{x\,(e) = x,\ y = 0\}$, $\{x\,(e) = x,\ y = 1\}$, ..., $\{x\,(e) = x,\ y = s\}$ sînt incompatibile între ele deci :

$$\{x\,(e) = x\} = \bigcup_{h=1}^{s} [x\,(e) = x,\ y = h]$$

și în consecință :

$$P\,[x\,(e) = x] = \sum_{h=1}^{s} P\,[x\,(e) = x,\ y = h]$$

sau :

$$f_1\,(x) = \sum_{y=0}^{s} f\,(x,\ y) \qquad (12_1)$$

Această formulă se putea deduce, de altfel, direct scriind că punctul M (x, y) se află pe dreapta x (e) = x, adică însumînd funcția f (x, y) pentru y = 0, 1, 2, ..., s.

În mod analog obținem distribuția marginală în raport cu y :

$$f_2(y) = \sum_{x=0}^{r} f(x, y) \qquad (12'_1)$$

Așa, de pildă, în exemplul 2°, distribuția marginală în raport cu x va fi :

$$f(x) = \sum_{y=0}^{3} f(x, y) = \sum_{y=0}^{3} \frac{C_5^x \cdot C_3^y C_6^{7-x-y}}{C_{14}^7} = \frac{C_5^x}{C_{14}^7} \sum_{y=0}^{3} C_3^y C_6^{7-x-y}$$

În baza identității (9_1) putem însă scrie :

$$\sum_{y=0}^{3} C_3^y C_6^{7-x-y} = C_9^{7-x}$$

așa încît avem densitatea marginală :

$$f_1(x) = \frac{C_5^x C_9^{7-x}}{C_{14}^7}, \quad 0 \leq x \leq 5$$

Această probabilitate se putea stabili direct modificînd enunțul problemei în felul următor : O urnă conține 14 bile din care 5 sînt albe. Se extrag în bloc 7 bile și se cere probabilitatea ca x să fie albe. Aplicînd formula (2_2), capitolul III, obținem imediat densitatea de mai sus.

Densitatea marginală în raport cu y va fi :

$$f_2(y) = \sum_{x=0}^{5} f(x,y) = \frac{C_3^y C_{11}^{7-y}}{C_{14}^7}$$

Funcția de repartiție a distribuției bidimensionale este probabilitatea :

$$P[x(e) < x,\ y(e) < y] = F(x,y) \qquad (13)$$

Din punct de vedere geometric, ea reprezintă probabilitatea ca punctul aleator M (s, t) să se găsească în dreptunghiul format

de axele de coordonate și de paralele la aceste axe duse prin punctele de abscisa x și ordonata y.

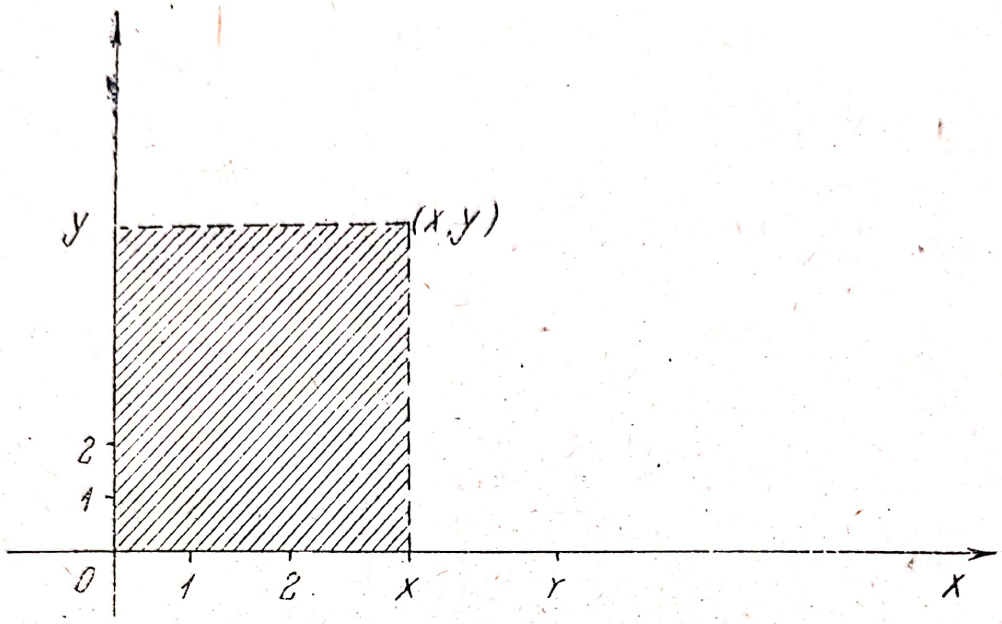

Fig. 22

Putem deci scrie relația :

$$F(x,y) = \sum_{s=0}^{x} \sum_{t=0}^{y} f(s, t) \qquad (13_1)$$

Noțiunile de densitate marginală și funcție de repartiție se extind fără modificare la o distribuție s-dimensională. Dacă distribuția cu s variabile aleatorii discrete $x_1, x_2, \ldots, x_s$, are densitatea $f(x_1, x_2, \ldots, x_s)$, atunci densitatea marginală în raport cu x_k de exemplu, va fi :

$$f_k(x_k) = \sum_{x_1=0}^{r_1} \cdots \sum_{x_{k-1}=0}^{r_{k-1}} \sum_{x_{k+1}=0}^{r_{k+1}} \cdots \sum_{x_s=0}^{r_s} f(x_1, x_2, \ldots, x_s) \qquad (14)$$

adică *se însumează funcția de densitate în raport cu toate celelalte variabile, afară de cea care ne interesează.*

Densitatea marginală în raport cu x_1 și x_2 va fi :

$$f_{12}(x_1, x_2) = \sum_{x_3=0}^{r_3} \sum_{x_4=0}^{r_4} \cdots \sum_{x_s=0}^{r_s} f(x_1, x_2, \ldots, x_s) \qquad (14_1)$$

Funcția de repartiție a distribuției de mai sus se definește în mod analog ca și pentru cea bidimensională :

$$F(x_1, x_2, \ldots, x_s) = \sum_{t_1=0}^{x_1} \sum_{t_2=0}^{x_2} \cdots \sum_{t_s=0}^{x_s} f(t_1, t_2, \ldots, t_s) \qquad (15)$$

Această funcție joacă un rol deosebit pentru distribuțiile continue, de aceea ne vom ocupa de ea mai pe larg în capitolul următor.

Probleme

1. Să se determine densitatea de probabilitate pentru numărul ce apare aruncînd un zar.
2. Care este densitatea pentru suma celor două numere care apar aruncînd două zaruri?
3. Se aruncă 10 zaruri. Să se determine densitatea numărului de apariții ale fețelor 3 și 5.
4. O urnă conține 7 bile albe și 4 negre. Se extrag succesiv 5 bile fără a fi puse înapoi în urnă. Să se calculeze densitatea numărului de bile negre.
5. Se aruncă trei monede de n ori. Să se determine densitatea pentru x, numărul de 5n în care nu apare nici o stemă, y numărul de ori în care stema apare o dată și z numărul de ori în care apar două steme.
6. Din piesele fabricate de o mașină $2^0/_0$ sînt rebuturi. Să se calculeze densitatea numărului de piese rebut dintr-un lot de 80 piese.
7. O urnă conține 5 bile albe și 10 negre. Se extrag bile, după schema bilei nerevenite, pînă se obțin 4 bile albe. Să se calculeze densitatea numărului total de bile extrase.
8. Pe baza unor probe numeroase s-a stabilit că 3% din semințele de mazăre nu germinează. Semințele se pun în pachete de 150 semințe și se garantează o germinație de 92%. Care este probabilitatea ca un pachet să nu corespundă garanției?
9. Un anumit volum mic de sînge conține în medie 24 globule roșii la persoanele normale. Să se calculeze probabilitatea ca o probă luată de la o persoană normală să conțină mai puțin de 19 globule roșii.
10. S-a stabilit că 0,001% din accidente se datoresc imprudenței pietonilor. Să se calculeze probabilitatea ca în cursul unui an să se întîmple mai mult de 6 accidente dacă sînt asigurate contra acestui risc 15 000 persoane.

Scanned with OKEN Scanner

11. În perioada de vîrf, o centrală telefonică primește în medie 700 chemări pe oră. Știind că ea poate da 25 legături pe minut să se calculeze probabilitatea ca ea să nu poată da legătura în cursul unui minut oarecare, folosind distribuția lui Poisson.

12. Se aruncă de 5 ori două zaruri. Să se calculeze probabilitatea ca numărul de cîte ori nu apare fața unu și numărul de cîte ori apar două fețe unu să fie fiecare mai mic decît 3.

Scanned with OKEN Scanner

Capitolul V

DISTRIBUȚII CONTINUE

§ 1. Distribuții unidimensionale

Să presupunem că variabila x reprezintă vîrsta unei persoane. Dacă w este limita maximă a vieții omenești, variabila x poate lua orice valoare din intervalul [o, w]. O asemenea variabilă se numește *variabila aleatorie continuă*, iar distribuția corespunzătoare se zice *distribuție de tip continuu*.

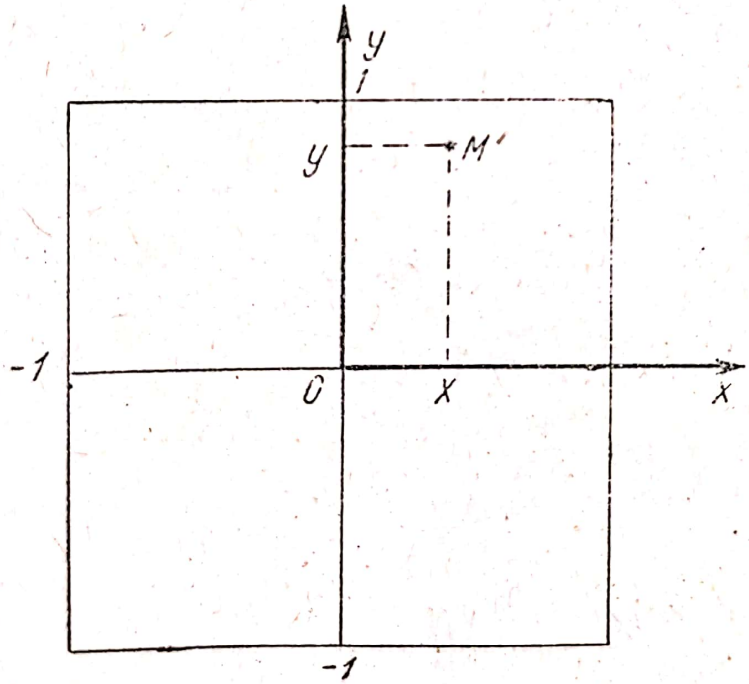

Fig. 23

Alt, exemplu. Să considerăm o țintă cu centrul 0, divizată în pătrate cu latura egală cu 1 (fig. 24). O armă a fost fixată în

așa fel ca gloanțele să nimerească centrul țintei. Într-o serie de trageri însă, punctul 0 nu va fi decît arareori nimerit, deoarece în cursul tragerii intervin o serie de factori care deviază traiectoria glonțului. Depărtarea x a punctului unde ținta a fost nimerită de verticala centrului este o variabilă aleatorie continuă ce poate lua orice valoare din intervalul $[-1,1]$. Am văzut că în cazul unei distribuții discrete variabila x lua un număr finit de valori sau o infinitate numărabilă. Probabilitatea ca variabila x să ia o valoare anumită x, pe care am notat-o prin f (x), avea în general o valoare diferită de zero și am reprezentat-o geometric prin niște segmente verticale.

În cazul unei variabile x de tip continuu nu mai putem lucra cu anumite valori individuale ale ei, ci cu intervale de valori de pe axa 0 x. În adevăr, experiența ne dovedește că într-o serie de trageri la țintă (vezi fig. 24), oricît de mare ar fi ea, practic este imposibil ca glonțul să nimerească în punctul de abscisă x (mai departe vom demonstra în mod riguros că probabilitatea acestui eveniment este egală cu zero). Tot experiența ne dovedește însă că într-o serie numeroasă de trageri putem obține lovituri în țintă a căror abscisă să aibă o valoare cuprinsă în intervalul $(-1, x)$. Probabilitatea unui asemenea eveniment este tocmai funcția de repartiție a variabilei x (e) definită în capitolul precedent :

$$F(x) = P[x(e) < x] \tag{1}$$

Funcția F (x) este în general o funcție continuă de x și satisface relațiile evidente :

$$F(a) = 0 \; F(b) = 1 \tag{1$_1$}$$

unde a și b sînt respectiv valoarea cea mai mică și cea mai mare pe care le poate lua x. Aceste relații le vom scrie mai departe sub o formă mai generală. În baza teoremei de adunare a probabilităților, rezultă că funcția de repartiție este *nedescrescătoare*, adică pentru oricare două valori x′ și $x'' > x'$ din intervalul (a, b) putem scrie :

$$\frac{F(x'') - F(x')}{x'' - x'} \geq 0 \tag{1$_2$}$$

Dacă F (x) este derivabilă vom nota derivata sa

$$F'(x) = f(x) \tag{1$_3$}$$

de unde deducem :

$$F(x) = \int_a^x f(x)\,dx \tag{1$_4$}$$

Funcția f (x) se numește *densitatea distribuției*. Deoarece F (x) este nedescrescătoare avem evident f (x) $\geqslant$ 0 și în baza condiției a doua din (1_1) putem scrie relația :

$$\int_a^b f(x)\,dx = 1 \qquad (1_5)$$

Din cele expuse pînă acum rezultă că o distribuție continuă este dată fie prin funcția sa de repartiție F (x), fie prin densitatea sa f (x).

Așadar o distribuție de tip continuu este caracterizată prin densitatea de probabilitate f (x) dacă funcția F (x) este derivabilă cu proprietățile :

$$1^\circ\ f(x) \geq 0$$
$$2^\circ\ \int_{-\infty}^{+\infty} f(x)\,dx = 1 \qquad (1_6)$$

Dacă intervalul în care variază x este finit condiția 2° se înlocuiește prin (1_5). Funcția de repartiție a variabilei x este definită prin relația :

$$F(x) = \int_{-\infty}^{x} f(t)\,dt \qquad (1_7)$$

iar cînd intervalul de variație este finit se înlocuiește cu relația (1_4) considerînd F (x) = 0 pentru x $\leqslant$ a. Funcția F (x) are proprietățile :

1°. Este nedescrescătoare.

2°. F ($-\infty$) = 0, F ($+\infty$) = 1, iar cînd intervalul este finit folosim relațiile (1_1) considerînd F (x) = 0 pentru x $\leqslant$ a și F (x) = 1 pentru x $\geqslant$ b.

3°. Dacă distribuția este continuă F′ (x) = f (x).

Observație : Din relațiile (1_6) rezultă că orice funcție nenegativă într-un anumit interval, convenabil modificată, poate reprezenta densitatea unei distribuții continue.

Exemplu : Funcția 3 x^2 — 1 este pozitivă în intervalul (1,3). Deoarece :

$$\int_1^3 (3x^2 - 1)\,dx = \left| x^3 - x \right|_1^3 = 27 - 3 - 1 + 1 = 24$$

rezultă că funcția :

$$f(x) = \begin{cases} 0 & \text{pentru } x \leq 1 \\ \frac{1}{24}(3x^2 - 1) & \text{pentru } 1 < x \leq 3 \\ 0 & \text{pentru } x > 3 \end{cases}$$

este o densitate de probabilitate. În adevăr, avem $f'(x) = \frac{1}{4} \times > 0$ pentru orice $x \in (1,3)$, deci $f(x)$ crește în acest interval, iar integrala :

$$\int_{-\infty}^{+\infty} f(x)\, dx = \int_{-\infty}^{1} 0\, dx + \int_{1}^{3} \frac{1}{24}(3x^2 - 1)\, dx + \int_{3}^{\infty} 0.\, dx = 1$$

Graficul densității $f(x) = \frac{1}{24}(3x^2 - 1)$ se numește *curba de distribuție* și una din problemele de bază ale statisticii matematice este determinarea acestei curbe din datele experimentale. Curba de distribuție pentru densitatea de mai sus are forma :

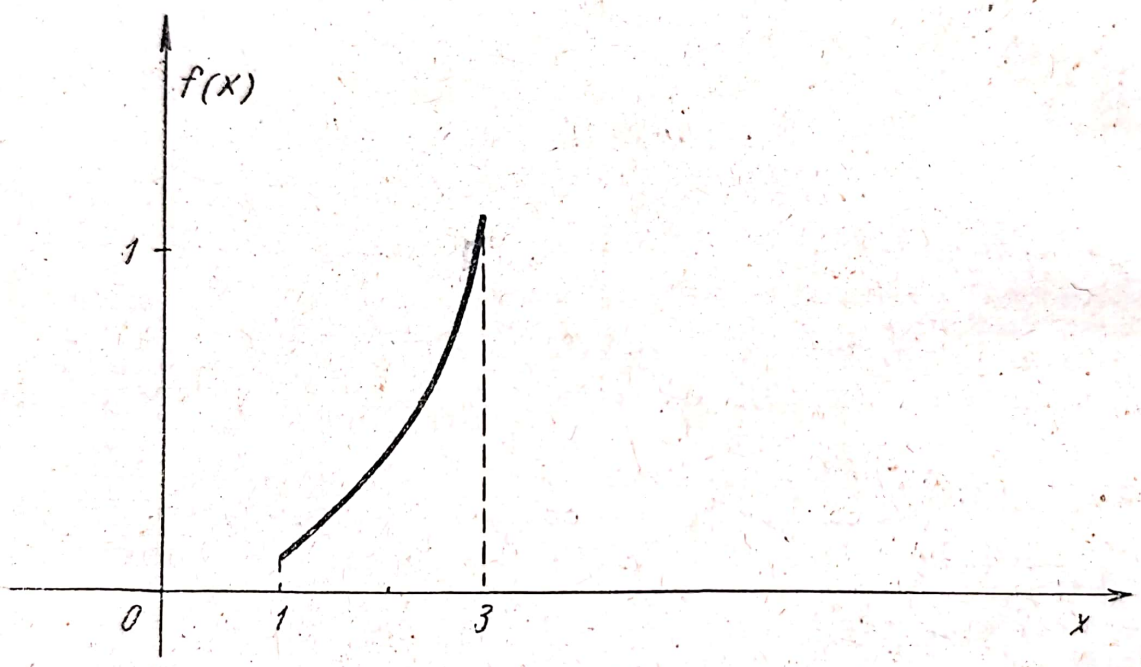

Fig. 24

Funcția de repartiție a distribuției va fi :

$$F(x) = \int_{1}^{x} f(x)\, dx = \begin{cases} 0, & \text{pentru } x \leq 1 \\ \frac{1}{24}(x^3 - x), & \text{pentru } 1 < x < 3 \\ 1 & \text{pentru } x \geq 3 \end{cases}$$

Graficul acestei funcții este redat mai jos :

Cunoașterea densității $f(x)$ ne permite rezolvarea unor probleme importante în legătură cu distribuția. Așa de pildă

probabilitatea ca variabila să ia o valoare dintr-un interval (x', x'') conținut în intervalul de definiție al distribuției se va obține integrînd funcția $f(x)$ în acest interval :

$$P(x' < x < x'') = \int_{x'}^{x''} f(x)\,dx \qquad (1_8)$$

(Reamintim că pentru distribuțiile discrete însumăm funcția de densitate în intervalul respectiv).

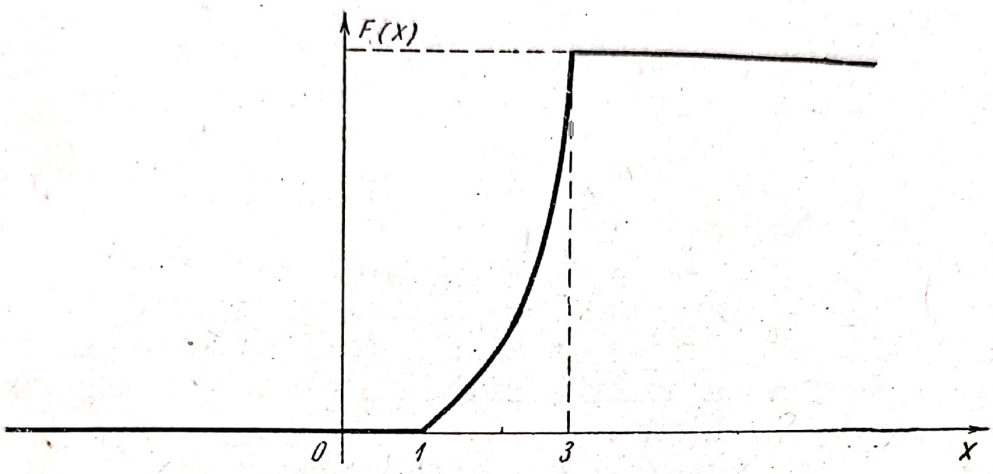

Fig. 25

Pentru densitatea :

$$f(x) = \frac{1}{24}(3x^2 - 1),\ 1 < x < 2$$

avem :

$$P(1 < x < 2) = \int_1^2 \frac{1}{24}(3x^2 - 1)\,dx = \frac{1}{24}\left| x^3 - x \right|_1^2 = \frac{1}{4}$$

Probabilitatea ca $x < \frac{3}{2}$ știind că x a luat o valoare mai mică decît 2 este probabilitatea condiționată

$$P_{x<2}\left(x < \frac{3}{2}\right) = \frac{P\left(x < \frac{3}{2}\right)}{P(x < 2)} = \frac{\int_1^{\frac{3}{2}} \frac{1}{24}(3x^2 - 1)\,dx}{\int_1^2 \frac{1}{24}(3x^2 - 1)\,dx} = \frac{\frac{15}{24 \cdot 8}}{\frac{6}{24}} = \frac{15}{48}$$

Deoarece putem scrie relația :

$$\int_{x'}^{x''} f(x)\,dx = \int_{-\infty}^{x''} f(x)\,dx - \int_{-\infty}^{x'} f(x)\,dx$$

ținînd seama de formula (1_7) putem scrie :

$$P(x' < x < x'') = F(x'') - F(x') \qquad (1_9)$$

adică *probabilitatea ca variabila să ia o valoare în intervalul* (x', x'') *se obține făcînd diferența valorilor funcției de repartiție pentru extremitățile intervalului.*

Observație. Punînd în (1_9) $x' = a - \varepsilon$, $x'' = a + \varepsilon$ găsim : $P(a - \varepsilon < x < a + \varepsilon) = F(a + \varepsilon) - F(a - \varepsilon)$.
După formula creșterilor finite avem : $F(a + \varepsilon) - F(a - \varepsilon) = 2\varepsilon F'(b)$ unde $b \in (a - \varepsilon, a + \varepsilon)$. Dacă $\varepsilon \to 0$, $x \to a$, deci :

$$P(x = a) = 0 \qquad (1_{10})$$

Probabilitatea ca x să ia valoarea a este deci egală cu zero, pentru orice a din intervalul de variație al variabilei aleatorii x. Acesta este motivul pentru care în studiul distribuțiilor continue nu lucrăm cu valorile individuale ale variabilei ca în cazul celor de tip discret.

Interpretarea geometrică a rezultatelor de mai sus. În virtutea proprietăților funcției de repartiție : este continuă, este nedescrescătoare și $F(-\infty) = 0$, $F(+\infty) = 1$, graficul acestei funcții admite ca asimptote axa ox, pentru $x = -\infty$ și dreapta $y = 1$, pentru $x = +\infty$. Deci curba $F(x)$ va avea forma de mai jos :

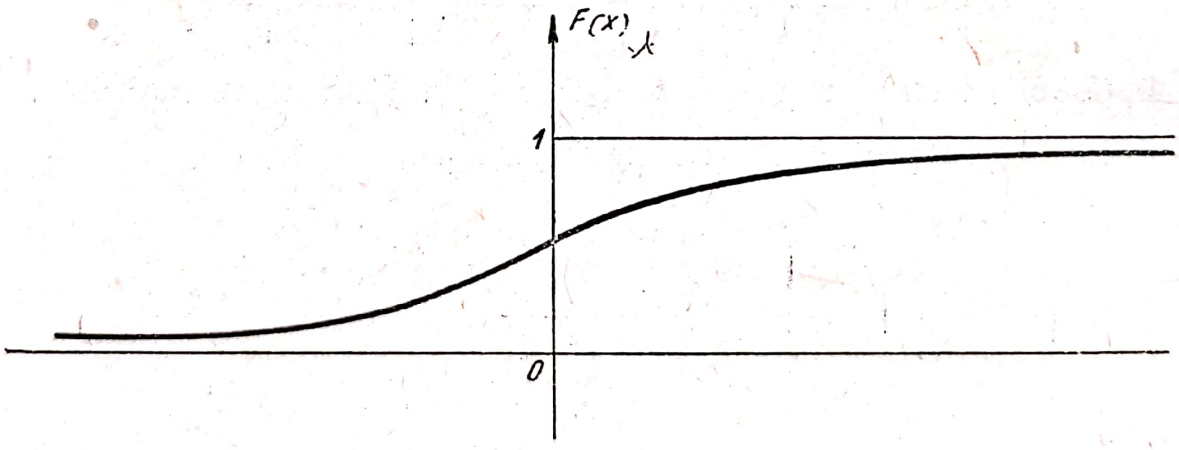

Fig. 23

În baza formulei (1_7), funcția de repartiție reprezintă aria mărginită de curba de distribuție f (x), de axa absciselor și

Scanned with OKEN Scanner

ordonata punctului x. În adevăr, integrala definită din partea a doua a relației (1_9) reprezintă această arie. În graficul de mai jos am haşurat această arie.

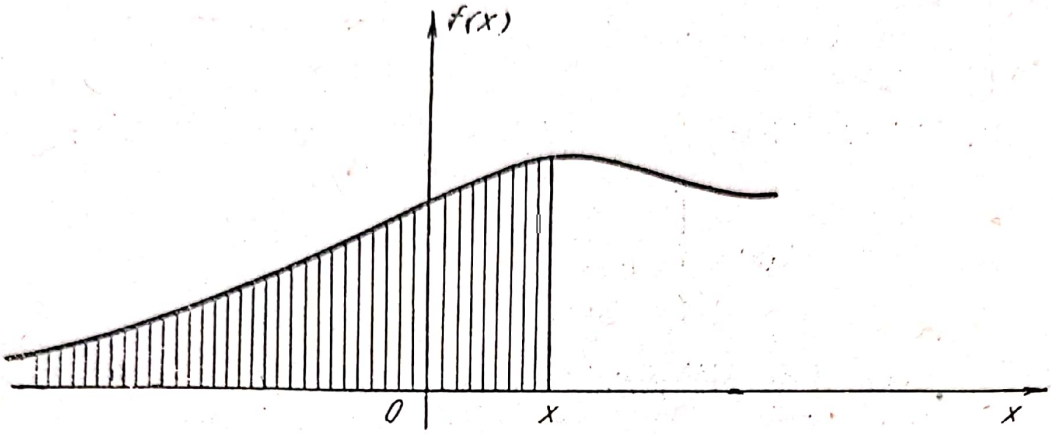

Fig. 27

Relația a doua din (1_6) arată că aria cuprinsă între curba de distribuție şi axa ox este egală cu 1, iar relația (1_8) arată că probabilitatea $P\ (x' < x < x'')$ este egală cu aria mărginită de curba de distribuție, axa ox şi ordonatele extremităților intervalului.

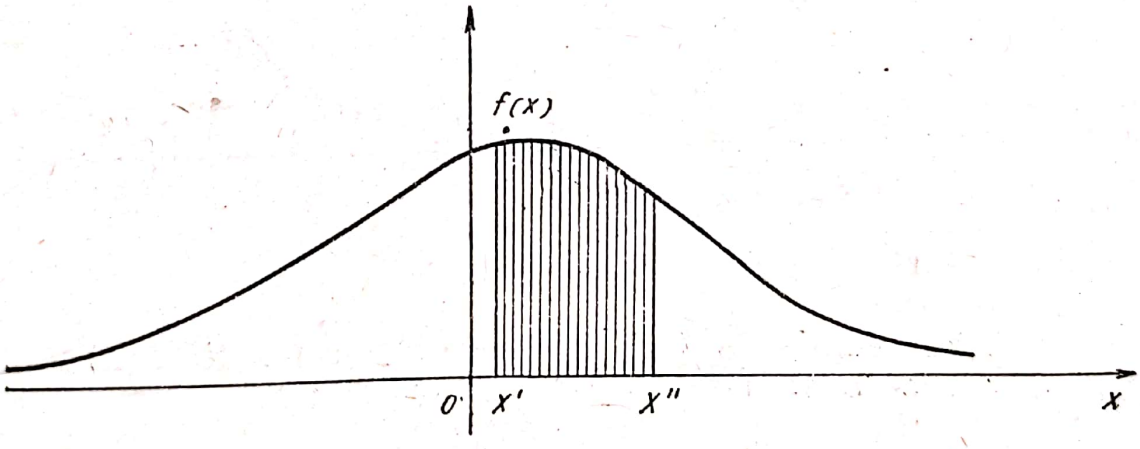

Fig. 28

Aria haşurată reprezintă probabilitatea

$$P\ (x' \leq x \leq x'') = \int_{x'}^{x''} f\ (x)\, dx = F\ (x'') - F\ (x') \qquad (1_{11})$$

Aria situată la stînga ordonatei punctului x' este, după interpretarea dată mai sus, egală cu funcția F (x'), iar cea situată la stînga ordonatei punctului x'' este F (x'').

Scanned with OKEN Scanner

Diferența lor, adică aria haşurată, este integrala definită (1_8).

Şi în acest caz putem folosi limbajul geometric de care ne-am folosit în capitolul precedent şi anume putem înlocui variabila x prin punctul aleator M (x).

Trecem acum la descrierea cîtorva distribuţii unidimensionale continue.

1°. *Distribuţia uniformă*, are densitatea :

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{pentru } a < x < b \\ 0, & \text{pentru celelalte valori} \end{cases} \qquad (2)$$

Graficul acestei densităţi este cel redat mai jos

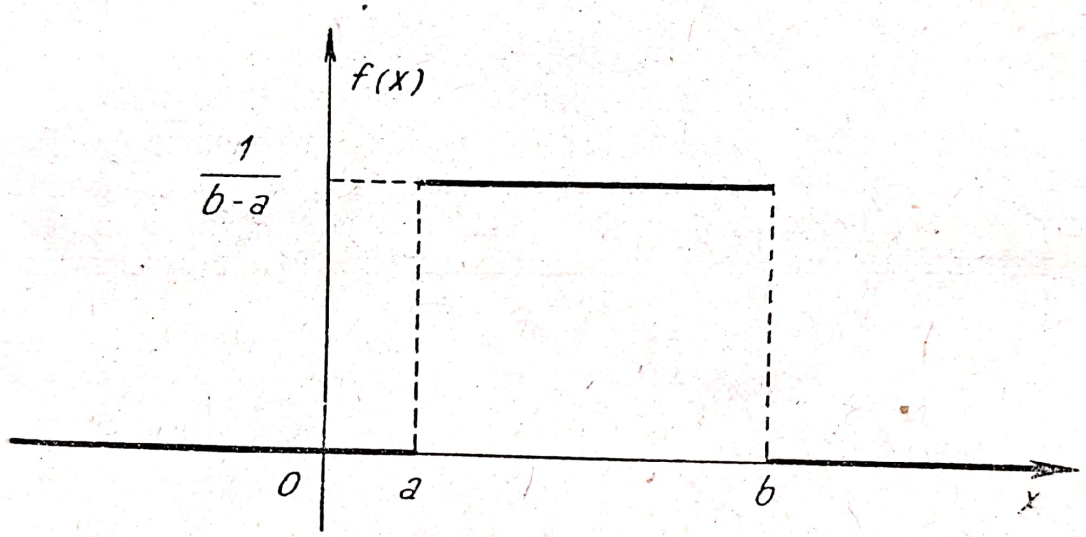

Fig. 29

Să verificăm condiţiile (1_8). Deoarece b > a rezultă f (x) > 0 pentru a < x < b. Pe de altă parte, integrala :

$$\int_{-\infty}^{+\infty} f(x)\,dx = \int_a^b \frac{1}{b-a}\,dx = \frac{1}{b-a}\,[x]_a^b = \frac{b-a}{b-a} = 1$$

Deci cele două condiţii (1_8) fiind verificate, funcţia (2) este o densitate de probabilitate. Funcţia de repartiţie a acestei distribuţii este :

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_a^b \frac{1}{b-a}\,dt = \frac{x-a}{b-a}$$

Prin urmare,

$$F(x) = \begin{cases} 0, & \text{pentru } x \leq a \\ \dfrac{x-a}{b-a}, & \text{pentru } a < x < b \\ 1, & \text{pentru } x \geq a \end{cases}$$

Graficul acestei funcții este următorul :

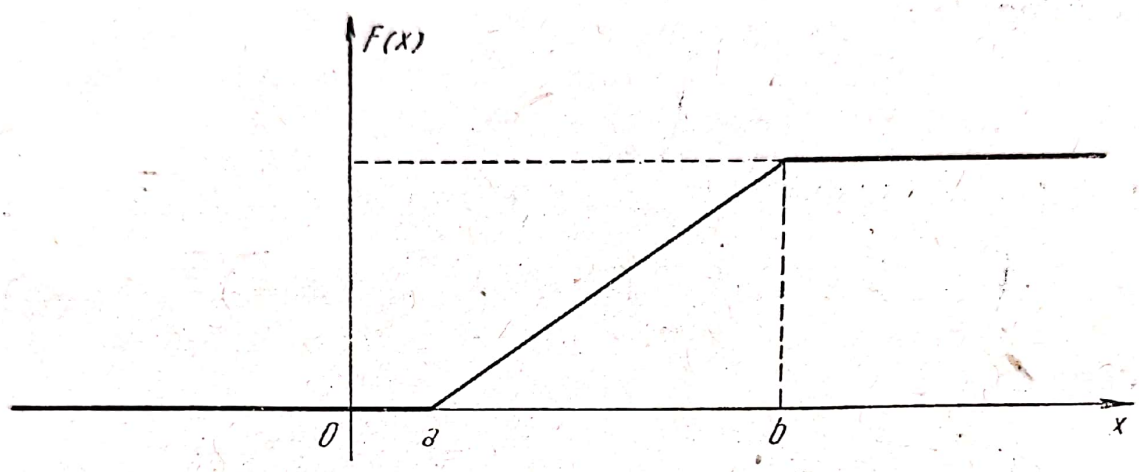

Fig. 30

2°. *Distribuția normală.* Funcția de densitate a acestei distribuții este următoarea :

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} \qquad -\infty < x < +\infty \qquad (3)$$

Distribuția normală formează o familie cu doi parametri m și σ. Uneori se notează funcția f (x) prin N (x ; m, σ). Se verifică ușor că f (x) $\geqslant$ 0 pentru orice x real. Pentru verificarea relației

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1$$

vom reaminti unele formule, utile în calculele care urmează. Integrala cunoscută :

$$\int_0^{\infty} e^{-x^2}\,dx = \frac{\sqrt{\pi}}{2} \qquad (\mathcal{S}_1)$$

devine prin substituția $x = \frac{t}{2}$, $dx = \frac{dt}{2}$

$$\int_0^{\infty} e^{-\frac{t^2}{2}} dt = \frac{\sqrt{2\pi}}{2} \qquad (3_1')$$

de unde deducem :

$$\frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{x^2}{2}} dx = \frac{1}{2} \qquad (3_2)$$

În integrala :

$$\int_{-\infty}^{0} e^{-\frac{x^2}{2}} dx$$

dacă facem schimbarea $x = -t$ $(t > 0)$, $dx = -dt$, obținem :

$$\int_{-\infty}^{0} e^{-x^2} dx = -\int_{+\infty}^{0} e^{-t^2} dt = \int_0^{+\infty} e^{-x^2} dx = \frac{\sqrt{\pi}}{2} \qquad (3_2')$$

de unde :

$$\int_{-\infty}^{+\infty} e^{-x^2} dx = \int_{-\infty}^{0} e^{-x^2} dx + \int_0^{\infty} e^{-x^2} dx = 2\int_0^{\infty} e^{-x^2} dx$$

Avem deci formula :

$$\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi} \qquad (3_3)$$

care prin substituția $t = x\sqrt{2}$, $dt = \sqrt{2}\, dx$ devine :

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = 1 \qquad (3_4)$$

Se numește funcția lui Laplace integrala

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_0^{t} e^{-\frac{x^2}{2}} dx \qquad (3_5)$$

Să revenim acum la integrala

$$\int_{-\infty}^{+\infty} f(x)\, dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx$$

Scanned with OKEN Scanner

şi să facem schimbarea de variabilă

$$\frac{x-m}{\sigma} = \sigma \; ; \; dx = \sigma \, dy \qquad (3_6)$$

Vom obţine :

$$\int_{-\infty}^{+\infty} f(x)\,dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}}\,dy = 1 \qquad \text{(relaţia } 3_4\text{)}$$

Aşadar f (x) este o densitate de probabilitate. Dacă $m = 0$, $\sigma = 1$, obţinem distribuţia *normală tip* :

$$f(x) = N(x;\, 0,1) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}, \quad -\infty < x < +\infty \qquad (3')$$

Distribuţia normală se zice *simetrică*, faţă de dreapta $x = m$, deoarece pentru orice a subzistă relaţia

$$f(m-a) = f(m+a) \qquad (4)$$

Prin urmare, curba (3) este simetrică faţă de dreapta $x = m$, iar curba (3′) este simetrică faţă de $x = 0$ (axa oy). Curbele cu acelaşi σ sînt identice ; ele coincid printr-o translaţie paralelă cu ox. Curbele cu acelaşi m sînt cu atît mai înalte cu cît σ este mai mic. Aceasta rezultă din (3), deoarece f (m $= \frac{1}{\sqrt{2\pi}\,\sigma}$, deci cu cît σ este mai mic cu atît f (m) este mai mare. În ce priveşte prima afirmaţie să considerăm două distribuţii cu acelaşi σ avînd parametrii m respectiv m′. Distribuţiile lor sînt deci :

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} \qquad f_1(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$$

Să formăm numerele f (m + a) şi f_1 (m′ + a).

Avem :

$$(m+a) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{1}{2}\left(\frac{a}{\sigma}\right)^2} \qquad f_1(m'+a) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{1}{2}\left(\frac{a}{\sigma}\right)^2}$$

de unde rezultă că $f(m + a) = f_1(m' + a)$. Aşadar punctele la distanţa a de axele respective de simetrie $x = m$ şi $x = m'$ au aceleaşi ordonate. Prin translaţia $|m - m'|$ cele două curbe de distribuţie coincid.

Redăm mai jos graficul curbei normale pentru $m = 3/2$ şi $\sigma = 1/2$, 1 şi 2

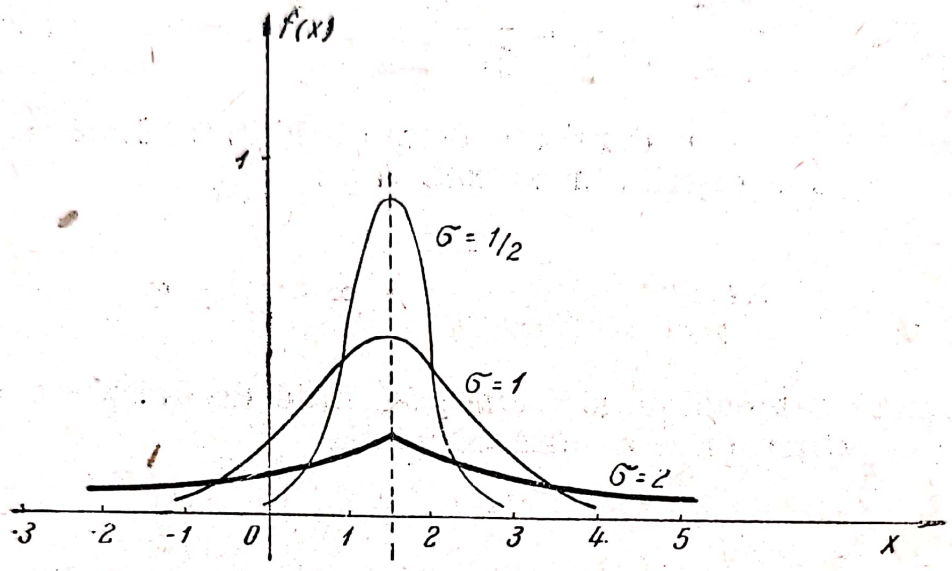

Fig. 31

Funcţia de repartiţie a distribuţiei normale este :

$$F(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{t-m}{\sigma}\right)^2} dt \qquad (4_1)$$

Făcînd substituţia :

$$\frac{t-m}{\sigma} = y\,;\, dt = \sigma\, dy$$

se găseşte :

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-m}{\sigma}} e^{-\frac{y^2}{2}} dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{y^2}{2}} dx +$$

$$+ \frac{1}{\sqrt{2\pi}} \int_{0}^{\frac{x-m}{\sigma}} e^{-\frac{y^2}{2}} dy$$

Ținînd seama de $(3'_2)$ în care facem $x = \frac{y}{\sqrt{2}}$ obținem :

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{y^2}{2}} dy = \frac{1}{2}$$

Ultima integrală este funcția lui Laplace (3_5) în care x se înlocuiește cu $\frac{x - m}{\sigma}$ deci :

$$F(x) = \frac{1}{2} + \Phi\left(\frac{x - m}{\sigma}\right). \qquad (4_2)$$

Probabilitatea ca x să ia o valoare din intervalul (x', x'') sau, ceea ce este același lucru, ca punctul aleatoriu M(x) să se găsească între punctele $M'(x')$ și $M''(x'')$ este

$$P(x' < x < x'') = \int_{x'}^{x''} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx = \frac{1}{\sqrt{2\pi}} \int_{\frac{x'-m}{\sigma}}^{\frac{x''-m}{\sigma}} e^{-\frac{t^2}{2}} dt \qquad (4_3)$$

sau încă folosind funcția de repartiție :

$$P(x' < x < x'') = \Phi\left(\frac{x''-m}{\sigma}\right) - \Phi\left(\frac{x'-m}{\sigma}\right). \qquad (4_4)$$

Distribuția normală joacă un rol de bază în statistica matematică. Ea a fost folosită pentru prima dată de către Moivre (1733) în studiul unei distribuții binomiale. Mai tîrziu, Gauss și Laplace ajung la legea normală în lucrările lor despre teoria erorilor de observație.

3°. *Distribuția Gama* este caracterizată prin densitatea de probabilitate

$$f(x) = \begin{cases} \dfrac{x^a\, e^{-\frac{x}{b}}}{\Gamma(a+1)\, b^{a+1}}, & \text{dacă } x > 0 \\ 0, & \text{dacă } x \leq 0 \end{cases} \qquad (5)$$

unde parametrii $a > -1$, $b > 0$.

Funcția $\Gamma(a+1)$ cunoscută sub numele de *funcția Gama* sau funcția euleriană de speța a doua este definită prin relația

$$\Gamma(a+1) = \int_0^\infty x^a \, e^{-x} \, dx \qquad (5_1)$$

Prin substituția $x = kt$, $dx = kdt$ integrala de mai sus se scrie :

$$\Gamma(a+1) = k^{a+1} \int_0^\infty t^a \, e^{-kt} \, dt$$

De unde deducem formula :

$$\int_0^\infty x^a \, e^{-kx} \, dx = \frac{\Gamma(a+1)}{k^{a+1}} \qquad (5_2)$$

În particular, pentru $k = \frac{1}{b}$ obținem :

$$\int_0^\infty x^a \, e^{-\frac{x}{b}} \, dx = b^{a+1} \, \Gamma(a+1) \qquad (5_3)$$

Ultimele două relații se reduc la (5_1) dacă $b = 1$.
Integrînd (5_1) prin părți în care considerăm $e^{-x} \, dx = d(-e^{-x})$ obținem relația :

$$\Gamma(a+1) = a \, \Gamma(a) \qquad (5_4)$$

Dacă a este un întreg pozitiv, putem folosi relația

$$\Gamma(a+1) = a \; ! \qquad (5_5)$$

În adevăr, din (5_4) avem : $\Gamma(a+1) = a\,\Gamma(a) = a(a-1)\,\Gamma(a-1) = a(a-1)(a-2)\,\Gamma(a-2) = a(a-1)(a-2) \ldots 2.1\,\Gamma(1)$ și cum

$$\Gamma(1) = \int_0^\infty e^{-x} \, dx = 1 \quad \text{rezultă } (5_5)$$

Să găsim valoarea $\Gamma\left(\frac{1}{2}\right)$ care se întîlnește des în aplicații. Vom avea :

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty x^{-\frac{1}{2}} \, e^{-x} \, dx = \sqrt{\pi}^{\,1} \qquad (5'_5)$$

[1] Am făcut substituția $x = \frac{1}{2} t^2$ și am folosit relația $(3'_1)$.

Din relațiile (5_1) și $(5'_5)$ deducem, pentru $a = \frac{1}{2}$

$$\Gamma\left(\frac{3}{2}\right) = \frac{\sqrt{\pi}}{4}, \quad \Gamma\left(\frac{5}{2}\right) = \frac{15}{8}\sqrt{\pi}$$

Să verificăm dacă funcția f(x) din (5) este o densitate de probabilitate. Se vede ușor că $f(x) > 0$ pentru $x > 0$. Din (5_3) deducem :

$$\int_{-\infty}^{+\infty} f(x)dx = \int_0^{\infty} \frac{x^a \, e^{-\frac{x}{b}}}{\Gamma(a+1)b^{a+1}} \, dx = \frac{1}{(a+1)\, b^{a+1}} \int_0^{\infty} x^a \, e^{-\frac{x}{b}} \, dx =$$

$$= \frac{\Gamma(a+1)\, b^{a+1}}{\Gamma(a+1)\, b^{a+1}} = 1$$

Funcția f(x) verifică cele două condiții ale densității de probabilitate, deci funcția (5) reprezintă o distribuție. Ea formează o familie cu doi parametri a și b. În graficul de mai jos am reprezentat curba de distribuție f(x), pentru $b = 1$ și $a = 0,1,3$.

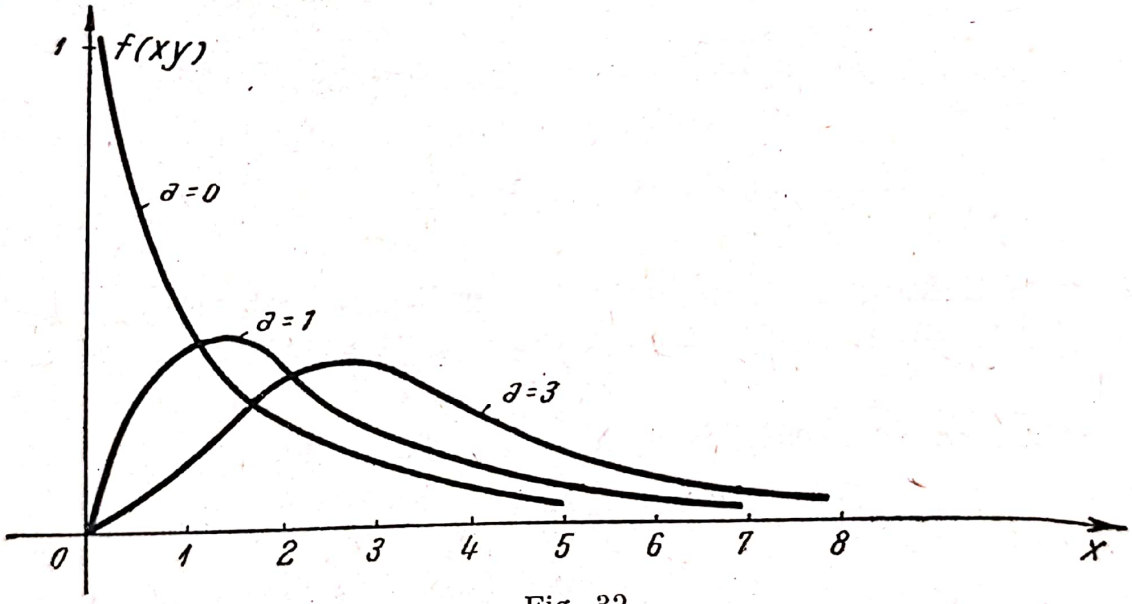

Fig. 32

Funcția de repartiție a distribuției Gama este :

$$F(x) = \begin{cases} \displaystyle\int_0^x \frac{1}{\Gamma(a+1)\, b^{a+1}} \, t^a \, e^{-\frac{t}{b}} \, dt, & (x > 0) \\ 0 & \text{pentru } x < 0 \end{cases}$$

numită și *funcția Gama incompletă*. Pentru calculul valorilor acestei funcții, K. Pearson a construit tabele speciale.

Scanned with OKEN Scanner

4°. *Distribuția Beta* are densitatea:

$$f(x) = \begin{cases} \dfrac{1}{B(a,b)}\, x^{a-1}(1-x)^{b-1}, & \text{pentru } 0 \leq x \leq 1 \\ 0, & \text{pentru celelalte valori} \end{cases} \tag{6}$$

unde $a > 0, b > 0$.

Funcția B (a, b) numită funcția Beta sau integrala eulerianã de speța I se definește prin relația:

$$B(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1}\, dx \tag{6_1}$$

Ea se exprimă cu ajutorul funcției Gama cu relația:

$$B(a, b) = \frac{\Gamma(a).\Gamma(b)}{\Gamma(a+b)} \tag{6_2}$$

Este evident că $f(x) > 0$ pentru $0 \leq x \leq 1$. Să calculăm integrala

$$\int_{-\infty}^{+\infty} f(x)\, dx = \int_0^1 \frac{1}{B(a,b).}\, x^{a-1}(1-x)^{b-1}\, dx =$$

$$= \frac{1}{B(a, b)} \int_0^1 x^{a-1}(1-x)^{b-1}\, dx.$$

În baza definiției (6_1) această integrală are valoarea 1, deci f(x) reprezintă o distribuție. Funcția f(x) se anulează pentru $x = 0$ și $x = 1$.

Funcția de repartiție a acestei distribuții este:

$$F(x) = \begin{cases} 0, \text{ pentru } x < 0 \\ \displaystyle\int_0^x \frac{1}{B(a,b)}\, t^{a-1}(1-t)^{b-1}\, dt, & \text{dacă } 0 < x < 1 \\ 1, \text{ pentru } x \geq 1 \end{cases}$$

Funcția F(x) se numește și funcția *Beta incompletă*.

5°. *Distribuția lui* χ^2 are densitatea

$$\text{unde } \xi = \chi^2 \quad f(\xi) = \frac{1}{2^{\frac{m}{2}}\Gamma\left(\frac{m}{2}\right)}\, \xi^{\frac{m}{2}-1}\, e^{-\frac{\xi}{2}}, \xi > 0 \tag{7}$$

Această distribuție, asupra căreia de altfel vom reveni mai tîrziu se obține din distribuția Gama pentru $a = \frac{m}{2} - 1$, $b = 2$. Ca exercițiu vom verifica însă direct condițiile (1) ale densității f(x).

În adevăr, este evident că f(x) este nenegativă pentru $x > 0$, iar integrala

$$\int_{-\infty}^{+\infty} f(\xi)\, d\xi = \frac{1}{\Gamma\left(\frac{m}{2}\right) \cdot 2^{\frac{m}{2}}} \int_0^{\infty} \xi^{\frac{m}{2}-1} e^{-\frac{\xi}{2}}\, d\xi$$

este egală cu 1 în baza relației (5_3) în care facem $a = \frac{m}{2} - 1$, $b = 2$.

6°. *Distribuția lui Student*. Variabila aleatorie t are o distribuție Student dacă:

$$f(t) = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi r}\,\Gamma\left(\frac{r}{2}\right)} \cdot \frac{1}{\left(1+\frac{t^2}{r}\right)^{\frac{r+1}{2}}}, \qquad -\infty < t < +\infty \tag{8}$$

7°. *Distribuția Z a lui R. A. Fisher* are densitatea

$$f(z) = \frac{2\, r^{\frac{r}{2}}, s^{\frac{s}{2}}}{B\left(\frac{r}{2}, \frac{s}{2}\right)} \cdot \frac{z^{\frac{r-2}{2}}}{\left(rz+s\right)^{\frac{r+s}{2}}} \tag{9}$$

8°. *Distribuția lui Cauchy* are densitatea

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}, \qquad -\infty < x < \infty \tag{10}$$

Funcția de repartiție a acestei distribuții este

$$F(x) = \int_{-\infty}^{x} \frac{1}{\pi} \frac{dx}{1+x^2} = \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} x$$

Se verifică imediat relația $F'(x) = f(x)$.

9°. *Distribuțiile lui Pearson*. O clasă generală de funcții de densitate este reprezentată prin familiile de soluții ale ecuației diferențiale

$$\frac{dy}{dx} = \frac{(x-a)\,y}{bx^2 + cx + d} \tag{11}$$

Scanned with OKEN Scanner

Această ecuație se obține scriind că panta dreptei care unește două puncte consecutive ale distribuției hipergeometrice este egală cu $\frac{dy}{dx}$. Soluțiile ecuației (11) au fost împărțite în 12 familii de curbe numite curbele lui Pearson și ele conțin ca pe niște cazuri particulare unele din distribuțiile descrise pînă acum. De pildă, distribuția Beta reprezintă curbele de tip I, distribuția Gama pe cele de tipul III, distribuția normală pe cele de tipul VII.

Familiile de curbe Pearson se obțin stabilind diverse relații între cei patru parametri, a, b, c, d ai ecuației diferențiale. De exemplu pentru $b = c = 0$ ecuația (11) se scrie:

$$\frac{dy}{y} = \frac{x+a}{d}\,dx \quad \text{a cărei soluție este} \quad y = k\,e^{\frac{(x+a)^2}{2d}}$$

Presupunînd $d < 0$ și k determinat în așa fel ca:

$$\int_{-\infty}^{+\infty} k\,e^{\frac{(x+a)^2}{2d}}\,dx = 1$$

obținem distribuția normală.

§ 2. Distribuții continue multidimensionale

Ca și în cazul distribuțiilor discrete ne vom ocupa mai întîi de distribuțiile bidimensionale. Un exemplu de asemenea distribuție se obține dacă în tragerile la țintă (fig. 24) vom considera și abaterile verticale ale punctului unde glonțul nimerește ținta, adică distanțele de la punct pînă la orizontala centrului O. Fiecare punct în care ținta a fost lovită este un punct aleatoriu M(x, y) ale cărui coordonate sînt variabilele aleatorii x și y.

Probabilitatea ca glonțul să nimerească ținta în regiunea situată la stînga verticalei punctului x și sub orizontala punctului y este o funcție de aceste variabile pe care o numim *funcția de repartiție a variabilelor x și y*. Această funcție este nedescrescătoare în raport cu fiecare variabilă în parte și continuă. Notînd-o prin F(x, y) avem:

$$F(x, y) = P[x(e) < x,\ y(e) < y] \qquad (12)$$

Scanned with OKEN Scanner

Dacă funcția admite derivate de ordinul al doilea vom nota :

$$\frac{\partial}{\partial x} \cdot \frac{\partial}{\partial y} F(x, y) = \frac{\partial^2 F}{\partial x \, \partial y} = f(x, y) \qquad (12_1)$$

de unde rezultă

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s, t) \, dt \, ds \qquad (12_2)$$

iar în cazul cînd cea mai mică valoare a lui x este a și cea mai mică valoare a lui y este b avem relația

$$F(x, y) = \int_{a}^{x} \int_{b}^{y} f(s, t) \, dt \, ds \qquad (12'_2)$$

În cazul exemplului de mai sus $a = b = -1$

Din cele arătate pînă acum rezultă $f(x, y) \geqslant 0$, deoarece $F(x, y)$ este nedescrescătoare. Pe de altă parte, cînd $x \to \infty$, $y \to \infty$ punctul aleator $M(x, y)$ se găsește sigur în tot planul, deci :

$$\left[F(x,y) \right]_{x=\infty, y=\infty} = F(\infty, \infty) = 1$$

sau încă :

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dy \, dx = 1$$

Funcția $f(x, y)$ *se numește densitatea de probabilitate a distribuției* de două variabile x și y. Ea se bucură de proprietățile :

$$f(x,y) \geqslant 0, \quad \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \, dy \, dx = 1 \qquad (12_3)$$

analoge cu relația (1_4) din cazul discret.

Funcția de repartiție are proprietățile :

1°. Este nedescrescătoare în raport cu fiecare variabilă în parte.

2°. $F(-\infty, y) = 0$, $F(x, -\infty) = 0$, adică este nulă dacă una din variabile tinde către $-\infty$.

3°. $F(\infty, \infty) = 1$.

4°. Dacă este continuă și derivabilă avem relația (12_1).

Așadar, o distribuție bidimensională este definită fie prin funcția de repartiție, fie prin densitatea sa de probabilitate.

Relațiile (12_2) și $(12'_2)$ ne arată că probabilitatea ca variabilele x și y să ia valori determinate într-o anumită regiune

Scanned with OKEN Scanner

se obține integrînd funcția f(x, y) în acea regiune. În particular, probabilitatea ca punctul aleator M(x, y) să se găsească în dreptunghiul $x = x'$, $x = x''$; $y = y'$, $y = y''$ unde $x'' > x'$, $y'' > y'$ este:

$$P(x' < x < x'', y' < y' < y'') = \int_{x'}^{x''} \int_{y'}^{y''} f(x, y)\, dy\, dx \qquad (13)$$

și în general

$$P\left[M(x, y) \in D \right] = \int \int_{D} f(x, y)\, dy\, dx \qquad (13')$$

unde D este un domeniu cuprins în domeniul de definiție al funcției f(x, y).

Interpretarea geometrică. Funcția de densitate f(x, y) reprezintă în spațiul cu trei dimensiuni o suprafață. Integrala (13)

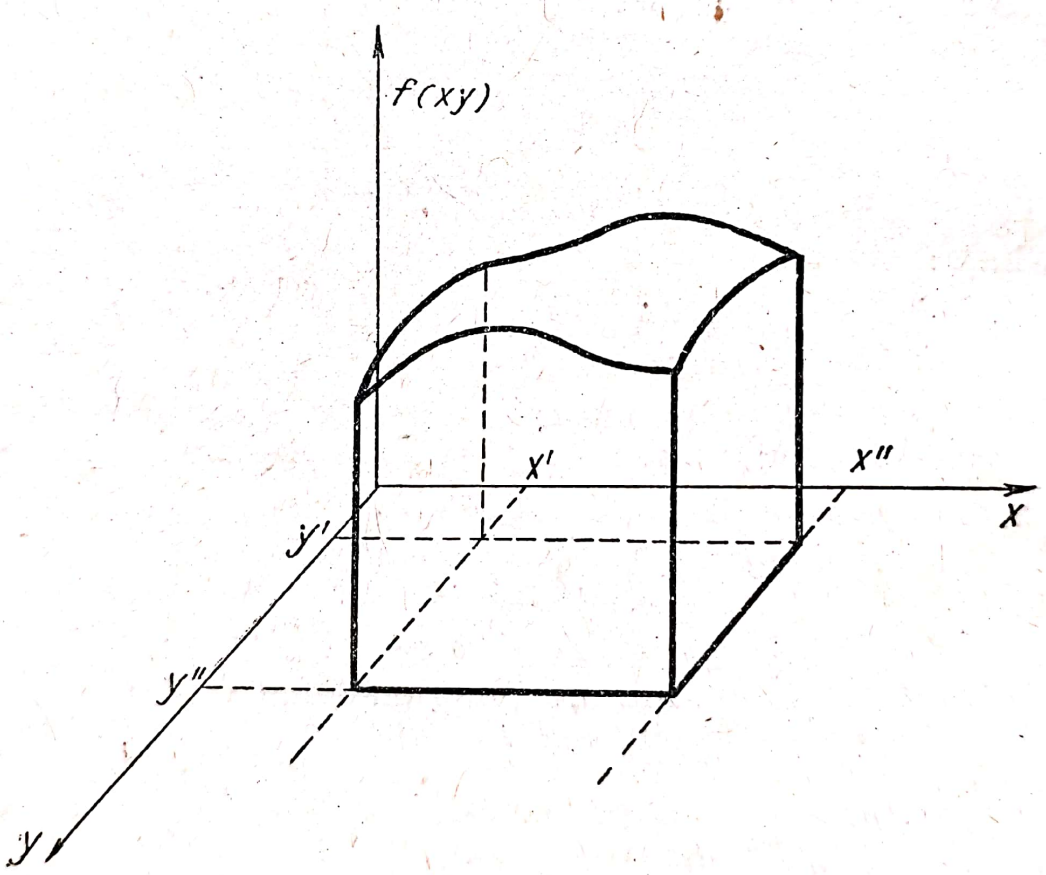

Fig. 33

reprezintă volumul paralelipipedului cu baza dreptunghiul din planul xy, fețele laterale paralele cu planele xz și yz respectiv și limitat la suprafata $z = f(x, y)$. În această interpretare,

funcția de repartiție a distribuției reprezintă volumul cuprins între suprafața f(x, y) și planul xy. Formula (13) se mai poate scrie :

$$P(x' < x < x'', y' < y < y'') = \int_{x'}^{x''} \int_{y'}^{y''} \frac{\partial}{\partial y} \frac{\partial}{\partial x} F(x, y)\, dy\, dx =$$

$$= \int_{x'}^{x''} \left[\frac{\partial}{\partial x} F(x, y)\right]_{y'}^{y''} dx = \int_{x'}^{x''} \frac{\partial}{\partial x} F(x, y'')\, dx - \int_{x'}^{x''} \frac{\partial}{\partial x} F(x, y')\, dx =$$

$$= \left[[F(x, y'')\right]_{x'}^{x''} - \left[F(x, y')\right]_{x'}^{x''}$$

De unde rezultă :

$$P(x' < x < x, y' < y < y'') = F(x'', y'') - F(x', y'') - F(x'', y') + F(x' y') \quad (13_1)$$

Interpretarea geometrică a acestei relații este următoarea : în planul xy, de pildă, F(x'', y'') este aria situată la stînga dreptei $x = x''$ și sub dreapta $y = y''$; F(x', y'') = aria la stînga dreptei $x = x'$ și sub $y = y''$ etc. ... ; suma algebrică a acestei arii este tocmai aria dreptunghiului x' x'' y' y''. Aceeași interpretare în spațiu : F(x'', y''), de pildă, reprezintă volumul cuprins între suprafața f(x, y) și planul xy, situat la stînga planului $x = x''$ și deasupra planului $y = y''$.

În baza condițiilor (12_3), orice funcție de două variabile nenegativă într-un anumit domeniu, convenabil modificată, poate fi considerată ca densitate de probabilitate a unei distribuții bidimensionale continue.

Exemplu. Funcția $f(x, y) = 6 - x - y$ este pozitivă de pildă în pătratul $0 < x < 2$; $2 < y < 4$ deoarece avem :

$$\int_0^2 \int_2^4 (6 - x - y)\, dy\, dx = 8$$

funcția următoare :

$$f(x, y) = \begin{cases} \frac{1}{8}(6 - x - y), & \text{pentru } 0 < x < 2;\ 2 < y < 4 \\ 0 & \text{pentru celelalte valori} \end{cases}$$

este o funcție de densitate, deoarece îndeplinește cele două condiții (12_3) cum este ușor să ne convingem. Vom rezolva cîteva probleme simple :

1°. Probabilitatea ca punctul aleator M(x, y) să apară în pătratul $x < 1$, $y < 3$ este :

$$P(x < 1, y < 3) = \int_{-\infty}^{1} \int_{-\infty}^{3} f(x, y)\, dy\, dx = \int_0^1 \int_2^3 \frac{1}{8}(6 - x - y)\, dy\, dx = \frac{3}{8}$$

2°. Probabilitatea $x + y < 3$ va fi :

$$P(x + y < 3) = \int_0^1 \int_2^{3-x} \frac{1}{8} (6 - x - y)\, dy\, dx =$$

$$= \int_2^3 \int_0^{3-y} \frac{1}{8} (6 - x - y)\, dx\, dy = 5/24$$

3°. Probabilitatea condiționată ca $x < 1$, dacă știm că $y < 3$ este :

$$P(x < 1) = P(x < 1/y < 3) = \frac{P(x < 1, y < 3)}{P(y < 3)}$$
$$y < 3$$

Probabilitatea $P(x < 1,\ y < 3)$ am calculat-o mai sus, iar probabilitatea $P(y < 3)$ se obține punînd condiția $y < 3$, indiferent de valorile lui x, adică :

$$P(y < 3) = \int_0^3 \int_2^3 \frac{1}{8} (6 - x - y)\, dy\, dx = 5/8$$

de unde rezultă : $P(x < 1/y < 3) = \frac{3}{8} : \frac{5}{8} = 3/5$

4°. Funcția de repartiție a acestei distribuții este :

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s,t)\, dt\, ds = \int_0^x \int_2^y \frac{1}{8} (6 - x - y)\, dy\, dx$$

Domeniul de definiție al distribuției, adică pătratul $0 < x < 2$, $2 < y < 4$, împarte planul în 9 regiuni :

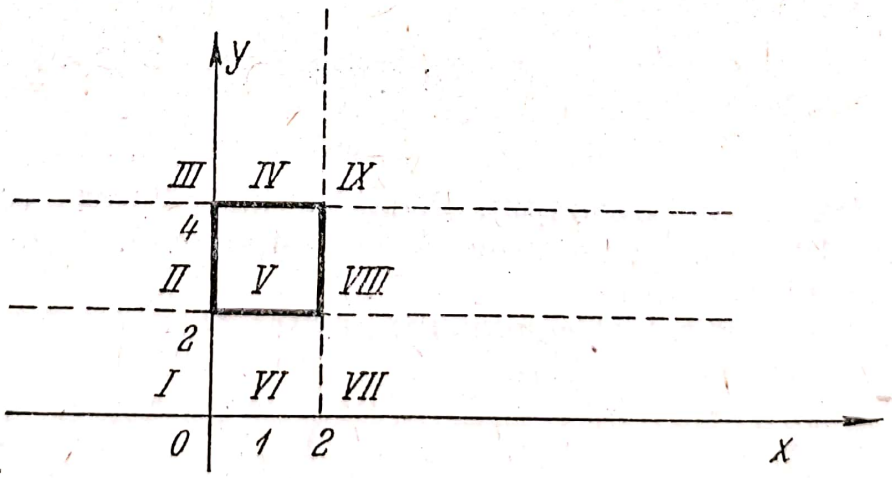

Fig. 34

În regiunile I, II, III, VI, și VII avem evident $F(x, y) = 0$. În regiunea V putem scrie :

$$F(x,y) = \int_0^x \int_2^y \frac{1}{8}(6 - s - t)\, dt\, ds = \frac{1}{8}\int_0^x \left[(6-s)(y-2) - \frac{y^2}{2} + 2\right] ds =$$

$$= \frac{1}{16} x (y-2)(10 - x - y)$$

Dacă punctul aleator se găsește în regiunea IV vom avea :

$$F(x,y) = \int_0^x \int_2^4 \frac{1}{8}(6 - s - t)\, dt\, ds = \frac{1}{8} \times (6-x)$$

Pentru regiunea VIII, $F(x, y) = \frac{1}{8}(y-2)(8-y)$ și în sfîrșit pentru IX obținem $F(x, y) = 1$. Aceste rezultate se pot concentra în următorul tablou

$$F(x,y) = \begin{cases} 0 \text{ dacă } x < 0 \text{ sau } y < 2 \\ \frac{1}{16} x (y-2)(10 - x - y), \quad 0 < x < 2 ; 2 < y < 4 \\ \frac{1}{8} x (6 - x) \qquad\qquad 0 < x < 2 ; y > 4 \\ \frac{1}{8}(y-2)(8-y), \qquad x > 2 ; 2 < y < 4 \\ 1, \qquad\qquad x > 2 ; y > 4 \end{cases}$$

Dacă facem în relația (13_1) : $x' = 0, x'' = 3 ; y' = 3, y'' = 4$ obținem probabilitatea : $P(0 < x < 1, \ 3 < y < 4) = F(1,4) - F(0,4) - F(1,3) + F(0,3) = \frac{1}{4}$

Distribuții marginale. Cu o distribuție bidimensională sînt asociate *distribuția marginală în raport cu x* și *distribuția marginală în raport cu* y.

Distribuția marginală în raport cu x este o funcție de x care integrată într-un anumit interval, fie (x', x'') ne va da probabilitatea ca punctul aleatoriu M (x, y) să se găsească pe ox în acest interval, oricare ar fi valoarea ordonatei y. În consecință, vom integra densitatea f(x, y) de la $-\infty$ la $+\infty$ în raport cu y și de la x' la x'' în raport cu x. Notînd f(x) probabilitatea marginală a distribuției, în raport cu x avem deci :

$$P(x' < x < x'') = \int_{x'}^{x''} \int_{-\infty}^{+\infty} f(x, y)\, dy\, dx \qquad (14)$$

Scanned with OKEN Scanner

Deoarece funcția de x,

$$f(x) = \int_{-\infty}^{+\infty} f(x, y)\, dy \qquad (14_1)$$

integrată de la x′ la x″ reprezintă tocmai probabilitatea $P(x' < x < x'')$, rezultă că (14_1) este distribuția, respectiv densitatea marginală în raport cu x. În mod analog, integrînd densitatea f(x, y) de la $-\infty$ la $+\infty$ în raport cu x, obținem densitatea marginală în raport cu y :

$$\varphi(y) = \int_{-\infty}^{+\infty} f(x, y)\, dx \qquad (14_2)$$

Exemplu : Pentru distribuția bidimensională continuă cu densitatea

$$f(x,y) = \frac{1}{8}(6 - x - y),\ 0 < x < 2;\ 2 < y < 4$$

avem următoarele distribuții marginale

$$f(x) = \begin{cases} \int_{-\infty}^{+\infty} f(x, y)\, dy = \int_{2}^{4} \frac{1}{8}(6 - x - y)\, dy = \frac{1}{4}(3 - x),\ 0 < x < 2 \\ 0, \text{ în celelalte cazuri} \end{cases}$$

$$\varphi(y) = \begin{cases} \int_{-\infty}^{+\infty} f(x, y)\, dx = \int_{0}^{2} \frac{1}{8}(6 - x - y)\, dx = \frac{1}{4}(5 - 2y),\ 2 < y < 4 \\ 0, \text{ în celelalte cazuri.} \end{cases}$$

Funcția de repartiție marginală se obține imediat din funcția de repartiție a distribuției bidimensionale după aceeași regulă ca mai sus. Astfel, notînd prin $F_1(x)$ și $\Phi_1(y)$ aceste funcții, avem :

$$F_1(x) = \int_{-\infty}^{x} \int_{-\infty}^{+\infty} f(x, y)\, dy\, dx = \int_{-\infty}^{x} f(s)\, ds \qquad (14_3)$$

$$\Phi_1(y) = \int_{-\infty}^{y} \varphi(t)\, dt \qquad (14_4)$$

Observație. Relațiile (14_3) și (14_4) ne arată că putem obține funcțiile de repartiție marginală calculînd funcțiile de repartiție ale distribuțiilor unidimensionale cu densitățile f(x) și respectiv φ(y) date prin relațiile (14_1) și (14_2).

Pentru distribuția particulară din exemplul precedent vom avea :

$$F_1(x) = \begin{cases} \int_{-\infty}^{x} f(s)\, ds = \int_{0}^{x} \frac{1}{4}(6 - s)\, ds = \frac{1}{8} x (6 - x),\ 0 < x < 2 \\ 1, \qquad x \geqslant 2 \end{cases}$$

Distribuții condiționate. Să presupunem că într-o anumită probă s-a determinat valoarea lui y nu și a lui x. Ne propunem să determinăm o funcție f(x) care să dea densitatea lui x condiționată de valoarea cunoscută a lui y. Cu alte cuvinte, să determinăm probabilitatea $P_y(x' < x < x'')$ pentru orice valori ale lui x' și x''.

Vom avea evident

$$P_y(x' < x < x'') = \int_{x'}^{x''} f_y(x)\,dx \tag{15}$$

Pentru calculul acestei probabilități considerăm intervalul $(y - a, y + a)$ în jurul punctului y și putem scrie formula probabilității condiționate astfel:

$$P_{y-a<t<y+a}(x' < x < x'') = \frac{P(x' < x < x'', y - a < t < y + a)}{P(y - a < t < y + a)}$$

sau încă:

$$P_{y-a<t<y+a}(x' < x < x'') = \frac{\int_{x'}^{x''}\int_{y-a}^{y+a} f_y(x, t)\,dt\,dx}{\int_{y-a}^{y+a}\int_{-\infty}^{+\infty} f(x, t)\,dx\,dt} \tag{15$_1$}$$

Deoarece $\int_{-\infty}^{+\infty} f(x, t)dx = \varphi(t)$ = densitatea marginală în raport cu t, numitorul se poate scrie:

$$\int_{y-a}^{y+a}\int_{-\infty}^{+\infty} f(x, t)\,dx\,dt = \int_{y-a}^{y+a} \varphi(t)\,dt = 2a\,\varphi(\xi)$$

unde $\xi \in (y - a, y + a)$.

Avem de asemenea:

$$\int_{y-a}^{y+a} f(x, t)\,dt = 2a\,f(x, \eta)$$

unde $\eta \in (y - a, y + a)$. Numărătorul se scrie atunci

$$\int_{x'}^{x''}\int_{y-a}^{y+a} f(x, t)\,dt\,dx = 2a\int_{x'}^{x''} f(x, \eta)\,dx$$

Scanned with OKEN Scanner

Probabilitatea (15_1) devine :

$$P\underset{y-a<t<y+a}{(x'<x<x'')} = \frac{\int_{x'}^{x''} f(x, \eta)\, dx}{\varphi(\xi)} \qquad (15_2)$$

Dacă $a \to 0$, atunci ξ, η și $t \to y$ așa încît obținem probabilitatea

$$P_y(x' < x < x'') = \frac{\int_{x'}^{x''} f(x, y)\, dx}{\varphi(y)} = \int_{x'}^{x''} \frac{f(x, y)}{\varphi(y)}\, dx$$

Această formulă fiind valabilă oricare ar fi x' și x'' rezultă, ținînd seama de relația (15) :

$$f_y(x) = \frac{f(x, y)}{\varphi(y)} \qquad (15_3)$$

În mod analog, probabilitatea lui y condiționată de x va fi :

$$\varphi_x(y) = \frac{f(x, y)}{f(x)} \qquad (15'_3)$$

Exemplu. Pentru distribuția cu densitatea

$$f(x, y) = \frac{1}{8}(6 - x - y), \quad 0 < x < 2;\ 2 < y < 4$$

avem :

$$\varphi_x(y) = \frac{f(x, y)}{f(x)} = \frac{6 - x - y}{2\,(3 - x)}, \quad \text{cînd } 2 < y < 4$$

deoarece $f(x)$ densitatea marginală în raport cu x este $\frac{1}{4}(3 - x)$.

Extinderea rezultatelor precedente la distribuțiile s — dimensionale. Noțiunile introduse pentru distribuțiile bidimensionale se extind fără nici o modificare și la distribuțiile cu mai mult de două dimensiuni.

În general, orice funcție $f(x_1, x_2, \ldots, x_s)$ poate fi considerată ca densitatea de probabilitate a unei distribuții cu s variabile aleatorii $x_1, x_2, \ldots, x_s$, dacă : $f(x_1, x_2, x_3 \ldots, x_s) \geqslant 0$

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, x_2, \ldots x_s)\, dx_1\, dx_2 \ldots dx_s = 1$$

Probabilitatea ca punctul aleator M $(x_1, x_2, \ldots, x_s)$ să apară într-un domeniu oarecare D din spațiul s-dimensional se obține integrînd funcția de densitate $f(x_1, x_2, \ldots, x_s)$ în acest domeniu :

$$P\left\{ M(x_1, x_2, \ldots x_s) \in D \right\} = \int\int \ldots \int_D f(t_1, t_2, \ldots t_s)\, dt_1\, dt_2 \ldots dt_s \qquad (16_1)$$

Funcția de repartiție a acestei distribuții este

$$F(x_1, x_2, \ldots, x_s) = P(x_1(e) < x_1, x_2(e) < x_2, \ldots, x_s(e) < x_s)$$

și se calculează cu formula

$$F(x_1, x_2, \ldots, x_s) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \ldots \int_{-\infty}^{x_s} f(t_1, t_2, \ldots, t_s)\, dt_1\, dt_2 \ldots dt_s \qquad (16_2)$$

Funcția $F(x_1, x_2, \ldots, x_s)$ are proprietățile :

1°. Este nedescrescătoare în raport cu fiecare variabilă în parte ;

2°. $F(x_1, x_2, \ldots x_{k-1}, -\infty, x_{k+1}, \ldots, x_s) = 0$, adică dacă una din variabile tinde către $-\infty$, funcția de repartiție este nulă.

3°. $F(+\infty, +\infty, \ldots + \infty) = 1$

4°. $$\frac{\partial^s}{\partial x_1, \partial x_2 \ldots \partial x_s} F(x_1, x_2, \ldots x_s) = f(x_1, x_2, \ldots x_s)$$

pentru distribuțiile continue.

Distribuția marginală în raport cu x_k, de pildă, se obține integrînd $f(x_1, x_2, \ldots, x_s)$ în raport cu $x_1 x_2, \ldots, x_{k-1}, x_{k+1}, \ldots, x_s$, între limitele $-\infty$ și $+\infty$. Mai general, distribuția marginală în raport cu o submulțime de variabile aleatorii se obține integrînd funcția de densitate în raport cu variabilele care nu ne interesează între $-\infty$ și $+\infty$.

Funcția de repartiție a distribuției marginale se obține din funcția de repartiție a distribuției făcînd să tindă către $+\infty$ variabilele care nu ne interesează.

Notații. Distribuția marginală în raport cu x_k o notăm $f_k(x_k)$, iar distribuția marginală în raport cu $x_k, x_h, \ldots, x_r$ o notăm prin $f(x_k, x_h, \ldots x_r)$. Funcția de repartiție a distribu-
$k, h \ldots r$

ției marginale în raport cu x_k o notăm prin $F_k(x_k)$, iar a distribuției marginale în raport cu $x_k, x_h, \ldots, x_r$ va fi $F(x_k, x_h, \ldots x_r)$.
$k,h \ldots r.$

Vom avea deci formula :

$$f_k(x_k) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(x_1, x_2, \dots x_k)\, dx_1\, dx_2 \dots dx_{k-1}\, dx_{k+1} \dots d_s \qquad (16_4)$$

Densitatea marginală în raport cu x_1, x_2, x_3, de pildă, este

$$f_{1\,2\,3}(x_1, x_2, x_3) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(x_1, x_2 \dots x_s)\, dx_4\, dx_5 \dots dx_s \qquad (16_5)$$

Funcția de repartiție marginală, în raport cu x_k va fi

$$F_k(x_k) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{x_k} \dots \int_{-\infty}^{+\infty} f(x_1, \dots x_2, \dots, x_s)\, dx_1\, dx_2 \dots dx_s =$$

$$= \int_{-\infty}^{x_k} f_k(x_k)\, dx_k \qquad (16_6)$$

În mod analog, vom putea scrie

$$F_{1\,2\,3}(x_1, x_2, x_3) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \int_{-\infty}^{x_3} f_{1,2,3}(t_1, t_2, t_3)\, dt_3\, dt_2 dt_1 \qquad (16_7)$$

Distribuția condiționată a variabilelor x_1, x_2, x_3, de pildă, de către variab lele x_4 și x_5 într-o distribuție cu variabilele x_1, x_2, x_3, x_4 și x_5 cu densitatea $f(x_1, x_2, x_3, x_4, x_5)$ va fi

$$f(x_1, x_2, x_3) = \frac{f(x_1, x_2, x_3, x_4\, x_5)}{f_{4\,5}(x_4, x_5)} \qquad (16_8)$$

$x_4, x_5)$

unde $f_{4\,5}(x_4, x_5)$ este distribuția marginală în raport cu x_4 și x_5, dată de formula :

$$f_{4,5}(x_4, x_5) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x_1, x_2, x_3, x_4, x_5)\, dx_1\, dx_2\, dx_3$$

Independența în sensul calculului probabilităților. Să considerăm o distribuție bidimensională cu densitatea $f(x, y)$. Prin definiție variabila x este independentă de variabila y dacă subzistă relația

$$f_y(x) = g(x), \qquad (17)$$

Scanned with OKEN Scanner

Cu alte cuvinte probabilitatea condiționată de y nu depinde de această variabilă. Deoarece avem :

$$f_y(x) = \frac{f(x,y)}{\varphi(y)}, \qquad \text{(relația } 15_3\text{)}$$

deducem : $f(x, y) = g(x)\ \varphi(y)$
Integrînd această relație de la $-\infty$ la $+\infty$ în raport cu y obținem :

$$\int_{-\infty}^{+\infty} f(x,y)dy = g(x) \int_{-\infty}^{+\infty} \varphi(y)\, dy$$

Avem însă relațiile cunoscute

$$\int_{-\infty}^{+\infty} f(x, y)\, dy = f(x) \ = \text{densitatea marginală}$$

în raport cu x ;

$$\int_{-\infty}^{+\infty} \varphi(y)\, dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)\, dy\, dx = 1$$

De unde rezultă $g(x) = f(x)$ așa încît,

$$f(x_1, y) = f(x)\, \varphi(y) \qquad (17_1)$$

Prin urmare, dacă *variabilele x și y sînt independente, densitatea distribuției este egală cu produsul densităților distribuțiilor marginale în raport cu fiecare variabilă.*

În general vom avea condiția necesară

$$f(x_1, x_2, \ldots, x_s) = f_1(x_1)\ f_2(x_2) \ldots, f_s(x_s) \qquad (17_2)$$

pentru independența două cîte două a variabilelor $x_1, x_2, \ldots x_s$.

Condiția necesară și suficientă pentru independența globală a variabilelor $x_1, x_2, \ldots x_s$ este ca pentru orice submulțime $x_{i_1}, x_{i_2}, \ldots x_{i_r}$ să avem :

$$f_{(x_{i_{r+1}}, \ldots x_{i_s})}\left(x_{i_1}, x_{i_2}, \ldots, x_{i_r}\right) = f\left(x_{i_1}\right) \cdot f\left(x_{i_2}\right) \ldots f\left(x_{i_r}\right) \qquad (17_3)$$

unde $i_1, i_2, \ldots i_s$ reprezintă o permutare oarecare a numerelor 1, 2, ..., s.

Aplicație : Să considerăm o distribuție cu 4 dimensiuni și funcția

$$f(x_1, x_2, x_3, x_4) = \begin{cases} 16\, x_1 x_2 x_3 x_4, & 0 \leq x_k \leq 1 \\ 0, \text{ pentru celelalte valori} \end{cases}$$

Avem evident $f > o$. Integrala multiplă

$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} 16\ x_1 x_2 x_3 x_4\, dx_1\, dx_2\, dx_3\, dx_4 = \int_0^1 \dots \int_0^1 16 \prod_{k=1}^{4} x_k \prod_{k=1}^{4} dx_k \equiv 1^1$$

deci funcția $16 \prod_{x=1}^{4} x_k$ reprezintă o densitate de probabilitate.

Densitatea marginală a variabilelor x_2 și x_4 este :

$$f_{24}(x_2, x_4) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x_1, x_2, x_3, x_4)\, dx_1\, dx_3 =$$

$$= 16 \int_0^1 \int_0^1 x_1 x_2 x_3 x_4\, dx_1\, dx_3 = 16\, x_2 x_4 \int_0^1 \int_0^1 x_1 x_3\, dx_1\, dx_3 \pm 4\, x_2 x_4$$

Prin urmare,

$$f_{24}(x_2, x_4) = \begin{cases} 4\, x_2 x_4,\ 0 \leq x_2 \leq 1\ ;\ 0 \leq x_4 \leq 1 \\ 0, \text{ pentru celelalte valori.} \end{cases}$$

Densitatea condiționată pentru x_1 și x_3 va fi :

$$f_{x_2, x_4}(x_1, x_3) = \frac{f(x_1, x_2, x_3, x_4)}{f_{24}(x_2, x_4)} = \frac{16\, x_1 x_2 x_3 x_4}{4 x_2 x_4} = 4\, x_1 x_3$$

deci :

$$f_{x_2, x_4}(x_1, x_3) = \begin{cases} 4\, x_1 x_3,\ 0 \leq x \leq 1\ ;\ 0 \leq x_3 \leq 1 \\ 0, \text{ pentru celelalte valori} \end{cases}$$

Cum această funcție nu depinde de x_2 și x_4 submulțimea x_1, x_3 este independentă de x_2, x_4 .

Funcția de repartiție a distribuției date, este

$$F(x_1, x_2, x_3, x_4) = \int_0^{x_1} \dots \int_0^{x_4} 16\, t_1 t_2 t_3 t_4\, dt_1\, dt_2\, dt_3\, dt_4 =$$

$$= 16 \int_0^{x_1} t_1\, dt_1 \int_0^{x_2} t_2\, dt_2 \int_0^{x_3} t_3\, dt_3 \int_0^{x_4} t_4\, dt_4 =$$

$$= 16 \prod_{k=1}^{4} \frac{\pi_k^2}{2} = x_1^2 \cdot x_2^2 \cdot x_3^2 \cdot x_4^2$$

---

[1] În adevăr, $\int_0^1 \dots \int_0^1 16 \prod_{k=1}^{4} x_k \prod_{k=1}^{4} dx_k = 16 \int_0^1 x_1\, dx_1 \dots \int_0^1 x_4\, dx_4 =$

$$= 16 \left[\int_0^1 x_1\, dx_1 \right]^4 = 16 \left\{ \left| \frac{x_1^2}{2} \right|_0^1 \right\}^4 = 16 \cdot \frac{1}{2^4} = 1$$

Scanned with OKEN Scanner

Printre distribuțiile continue multidimensionale, menționăm distribuția normală de care ne vom ocupa mai de aproape într-unul din capitolele următoare.

Probleme

1. Funcția $f(x) = \frac{1}{12}(2x-3)$ pentru $2 \leqslant x \leqslant 5$ este o densitate de probabilitate.

2. Dacă $f(x) = 2x$ pentru $0 \leqslant x \leqslant 1$ și nulă pentru celelalte valori să se calculeze probabilitățile:

1°. $P(x > 1/2)$

2°. $P\left(\frac{1}{4} \leq x \leq \frac{1}{2}\right)$

3°. $P\left(x > \frac{3}{4}\right)$ știind că $x > ½$.

3. Variabila aleatorie x are densitatea

$$f(x) = 3x^2, \text{ pentru } 0 \leq x \leq 1$$
$$= 0, \text{ pentru celelalte valori}$$

Să se determine un număr n așa fel ca x să ia cu egală probabilitate valori mai mari sau mai mici decît n.

4. Admițînd că viața medie în ore a unui anumit tip de lămpi de radio are densitatea:

$$f(x) = 100\,x^2, \text{ dacă } x \geq 100$$
$$= 0, \text{ pentru } x < 100$$

să se calculeze:

1° Probabilitatea ca din 3 lămpi să nu fie nevoie să se înlocuiască niciuna în primele 150 ore de funcționare;

2° Probabilitatea ca toate lămpile să fie înlocuite în primele 150 ore.

5. O mașină fabrică axe al căror diametru x are densitatea $f(x) = C(x-0{,}24)^2(x-0{,}26)^2$, în cazul cînd $0{,}24 \leqslant x \leqslant 0{,}26$, sau egală cu zero pentru celelalte valori. C este o constantă care se determină prin condiția ca f(x) să reprezinte o densitate de probabilitate. Axele merg la rebut dacă diametrele lor diferă cu mai mult de 8‰ față de numărul 0,25. Să se determine proporția probabilă a axelor care vor fi rebutate.

Scanned with OKEN Scanner

6. Dacă $f(x, y) = 1$ pentru $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1$ și nulă pentru celelalte valori, să se determine probabilitățile :

1° $x \leq \frac{1}{2}$, $y < \frac{1}{2}$

2° $x + y < 1$

3° $x > 2y$

4° $x + y > 1$

5° $x > \frac{1}{3}$

6° $x^2 + y^2 < \frac{1}{4}$

7° $x = y$

8° $x > \frac{1}{2}$ dacă $y < \frac{1}{2}$

7. Se dă densitatea :

$$f(x, y) = \begin{cases} e^{-(x+y)}, \text{ dacă } x > 0, y > 0 \\ 0, \text{ pentru celelalte valori} \end{cases}$$

Să se calculeze :

1° $P(x > 1)$

2° $P a < x + y < b)$, dacă $0 < a < b$

3° $P(x < y/x < 2y)$

4° Densitatea marginală a lui x

5° $P(0 < x < 1/y = 2)$

8. Să se reprezinte grafic distribuția :

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot 0{,}25} e^{-\frac{1}{2}\left(\frac{x}{0{,}25}\right)^2}$$

9. Să se calculeze a așa fel ca pentru variabila normală x să avem : $P(m - a\sigma \leqslant x \leqslant m + a\sigma) = 0{,}95$.

Scanned with OKEN Scanner

Capitolul VI

VALORI MEDII. MOMENTE

§ 1. Definiția valorii probabile pentru variabilele aleatorii de tip discret

Fie x o variabilă aleatorie de tip discret definită pe mulțimea E_k, $1 \leqslant k \leqslant n$ de evenimente elementare ale unui cîmp [E, k, P] și densitatea sa de probabilitate

$$f(x) = P[x(e) = x] \tag{1}$$

cu proprietățile cunoscute

$$f(x) \geqq 0, \quad \Sigma f(x) = 1 \tag{1_1}$$

unde însumarea se face după toate valorile lui x.

Valoarea medie sau valoarea probabilă a variabilei x, pe care o notăm prin M (x), se definește prin relația

$$M(x) = \Sigma x f(x) \tag{2}$$

Uneori valoarea medie a variabilei x se mai notează și prin E (x) sau, mai simplu, prin $\bar{x}$.

Exemple: 1°. Variabila x cu densitatea uniformă $f(x) = 1/n$ ia valorile $x_1, x_2, \ldots, x_n$ deci:

$$M(x) = \sum_{k=1}^{n} x_k f(x_k) = \sum_{k=1}^{n} x_k \cdot \frac{1}{n} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n)$$

În cazul aruncării unui zar avem: $x_k = k$ $(1 \leqslant k \leqslant 6)$

$$M(x) = \frac{1}{6}(1 + 2 + \ldots + 6) = 3{,}5$$

Scanned with OKEN Scanner

2°. Pentru distribuția binomială cu densitatea $f(x) = C_n^x p^x q^{n-x}$ $(x = 0,1, 2,\ldots n)$ și $p + q = 1$ se găsește valoarea medie:

$$M(x) = \sum_{x=0}^{n} xf(x) = \sum_{x=0}^{n} x C_n^x p^x q^{n-x} = np \qquad (3)$$

În adevăr scriind dezvoltarea:

$$(pt + q)^n = \sum_{x=0}^{n} C_n^x p^x q^{n-x} t^x$$

obținem prin derivare în raport cu t:

$$np(pt + q)^{n-1} = \sum_{x=0}^{n} x C_n^x p^x q^{n-x} t^{x-1}$$

De unde, făcînd $t = 1$ se obține relația (3).

Am văzut la studiul distribuției binomiale că valoarea cea mai probabilă a numărului de cîte ori se produce evenimentul A în n probe este $m = np$. În baza rezultatului de mai sus deducem că valoarea cea mai probabilă este egală cu valoarea medie a distribuției.

3°. Pentru distribuția lui Poisson cu densitatea:

$$f(x) = e^{-m} \frac{m^x}{x!}, x = 0, 1, 2, \ldots$$

avem valoarea medie:

$$M(x) = \sum_{x=0}^{\infty} x f(x) = e^{-m} \sum_{x=0}^{\infty} \frac{x m^x}{x!} = e^{-m} . m \sum_{m=1}^{\infty} \frac{m^{x-1}}{(x-1)!}$$

Punînd $x - 1 = y$ avem:

$$\sum_{y=0}^{\infty} \frac{m^y}{y!} = e^m$$

deci $M(x) = e^{-m}. m. e^m$, de unde:

$$M(x) = m \qquad (3_1)$$

4°. Să considerăm distribuția hipergeometrică:

$$f(x) = \frac{C_a^x C_{m-a}^{n-x}}{C_m^n}, \quad x = 0, 1, 2, \ldots . n$$

Scanned with OKEN Scanner

Valoarea medie va fi :

$$M(x) = \sum_{x=0}^{n} x\, f(x) = \sum_{x=0}^{n} \frac{x\, C_a^x \cdot C_{m-a}^{n-x}}{C_m^n}$$

Am demonstrat însă relația :

$$\sum_{x=0}^{n} C_a^x C_{m-a}^{n-x} = C_m^n \qquad (\alpha)$$

Deoarece :

$$x\, C_a^x = \frac{a(a-1)(a-2)\ldots(a-x+1)}{(x-1)!} = a\, C_{a-1}^{x-1}$$

vom avea :

$$\sum_{x=0}^{n} x\, C_a^x C_{m-a}^{n-x} = a \sum_{x=1}^{n} C_{a-1}^{x-1} C_{m-a}^{n-x} = a\, C_{m-1}^{n-1}$$

în baza relației (α) în care se înlocuiește a cu $a-1$ și x cu $x-1$. În acest caz vom obține deci valoarea medie

$$M(x) = a\, \frac{C_{m-1}^{n-1}}{C_m^n} = a \cdot \frac{n}{m} \cdot = \frac{a}{m} \cdot n = np \qquad (3_2)$$

în care $p = \frac{a}{m}$ și reprezintă probabilitatea de realizare a evenimentului A într-o singură probă.

Dacă *variabila x este continuă*, valoarea sa medie se definește în mod analog, suma înlocuindu-se printr-o integrală. Vom avea deci :

$$M(x) = \int_{-\infty}^{+\infty} x\, f(x)\, dx \qquad (4)$$

dacă integrala din partea dreaptă are sens.

Cînd intervalul de variație este finit, integrala se consideră nulă pentru valorile variabilei în afara intervalului.

Exemple : 1°. Pentru distribuția uniformă

$$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, \text{ pentru celelalte valori} \end{cases}$$

obținem :

$$M(x) = \int_{-\infty}^{+\infty} x\, f(x) = \int_a^b x\, \frac{1}{b-a}\, dx = 1\ \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b$$

Scanned with OKEN Scanner

De unde:

$$M(x) = \frac{b+a}{2} \tag{4_1}$$

2°. Valoarea medie a distribuției normale va fi:

$$M(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} x\, e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx$$

Făcînd schimbarea $\frac{x-m}{\sigma} = t$, $dx = \sigma\, dt$ avem:

$$M(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} (m + \sigma t)\, e^{-\frac{t^2}{2}} \sigma dt =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (m + \sigma t)\, e^{-\frac{t^2}{2}} dt = \frac{m}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt + \frac{\sigma}{2} \int_{-\infty}^{+\infty} t\, e^{-\frac{t^2}{2}} dt$$

În baza relației (3_2), capitolul V avem: $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt = 1$

Pe de altă parte avem:

$$\frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t\, e^{-\frac{t^2}{2}} dt = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} d\left(-\frac{t^2}{2}\right) = -\frac{\sigma}{\sqrt{2\pi}} \left[e^{-\frac{t^2}{2}}\right]_{-\infty}^{+\infty} = 0$$

Rezultă deci valoarea medie a distribuției normale

$$M(x) = m \tag{4_2}$$

Prin urmare, *parametrul m este valoarea medie a distribuției.*

3°. Distribuția Gama are valoarea medie

$$M(x) = \int_{-\infty}^{+\infty} x \frac{x^a e^{-\frac{x}{b}}}{\Gamma(a+1)\, b^{a+1}} dx = \frac{1}{\Gamma(a+1)\, b^{a+1}} \int_0^{\infty} x^{a+1} e^{-\frac{x}{b}} dx$$

Din relația (5_3), capitolul V deducem:

$$\int_0^{\infty} x^{a+1} e^{-\frac{x}{b}} dx = b^{a+2} \Gamma(a+2) = b^{a+2} (a+1) \Gamma(a+1)$$

De unde rezultă valoarea medie

$$M(x) = b\,(a+1) \tag{4_3}$$

Scanned with OKEN Scanner

4°. Valoarea medie a distribuției Beta este următoarea:

$$M(x) = \int_{-\infty}^{+\infty} xf(x)\,dx = \int_0^1 x\,\frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}\,dx =$$

$$= \frac{1}{B(a,b)}\int_0^1 x^a(1-x)^{b-1}\,dx = \frac{B(a+1,b)}{B(a,b)}$$

După relația (6_3) capitolul V avem:

$$\frac{B(a+1,b)}{B(a,b)} = \frac{\Gamma(a+1)\cdot\Gamma(b)}{\Gamma(a+b+1)} \cdot \frac{\Gamma(a+b)}{\Gamma(a)\cdot\Gamma(b)}$$

De unde:

$$M(x) = \frac{a}{a+b} \qquad (4_4)$$

Alte valori medii. În statistica matematică se folosesc și alte valori medii care sînt îndeosebi utile pentru studiul distribuției continue. Le vom enumera pe cele mai importante dintre ele.

1°. *Valoarea modală sau modul* unei distribuții este valoarea variabilei x pentru care densitatea de probabilitate este maximă. Dacă funcția f(x) are mai multe maxime densitatea se numește *multimodală*; în cazul unui singur maxim distribuția se zice unimodală. Modul se notează de obicei prin $M_o(x)$. Valorile modale se obțin rezolvînd ecuația $f'(x) = 0$ și luînd acele rădăcini r pentru care $f''(r) < 0$.

Acestea sînt, după cum se știe, condițiile ca valorile $x = r$ să fie maximele funcției f(x).

Exemplu. Distribuția cu densitatea:

$$f(x) = \sqrt{\frac{x}{\pi}}\,e^{-x}, \qquad x \geq 0$$

este o distribuție Gama — $(a = \frac{1}{2}, b = 1)$, iar derivata sa

$$f'(x) = \frac{e^{-x}}{2\sqrt{\pi x}}(1 - 2x)$$

se anulează pentru $x = 1/2$.

Deoarece derivata $f''(x) = \frac{-e^{-x}}{4x\sqrt{\pi x}}(1 + 4x - 4x^2)$ este negativă pentru $x = 1/2$, rezultă $M_o(x) = \frac{1}{2}$

Scanned with OKEN Scanner

2°. *Cvantilele distribuției* se definesc prin relația :

$$F(x) = k \tag{5}$$

unde evident $0 \leqslant k \leqslant 1$, iar x este cvantila căutată; funcția F (x) reprezintă ca de obicei funcția de repartiție a distribuției.

În particular, pentru $k = 1/2$ se obține *mediana distribuției* care se notează M_e (x); deci mediana se calculează din relația

$$F(Me) = \int_{-\infty}^{M_e} f(x)\,dx = 1/2 \tag{5_1}$$

Din punct de vedere geometric, mediana este abscisa punctului a cărui ordonată împarte aria cuprinsă între curba f(x) și axa absciselor în două părți echivalente. De aici rezultă că pentru *distribuțiile simetrice*, cum este de pildă distribuția normală, *mediana este egală cu valoarea medie a distribuției.*

Exemplu. Fie distribuția cu densitatea :

$$f(x) = \frac{1}{12}(2x + 1) \qquad 0 < x < 3$$

Aplicînd formula (5_1) obținem :

$$\int_{-\infty}^{M_e} f(x)\,dx = \int_{0}^{M_e} \frac{1}{12}(2x + 1)\,dx = \left[\frac{1}{12}(x^2 + x)\right]_0^{M_e}$$

De unde ecuația $\frac{1}{12}(M_e^2 + M_e) = \frac{1}{2}$ adică : $M_e^2 + M_e - 6 = 0$

Rădăcinile acestei ecuații sînt $M_e' = -3$, $M_e'' = 2$. A doua rădăcină 2, fiind cuprinsă în intervalul (0, 3), este mediana căutată.

Observație. Din definiția (5) rezultă următoarea construcție grafică : Se construiește curba $y = F(x)$ și se intersectează cu dreapta $y = k$, în particular, pentru mediana cu dreapta $y = 1/2$; abscisa punctului de intersecție este cvantila (mediana) căutată. Dacă distribuția este bimodală de pildă, mediana nu este unic determinată prin relația (5_1). În acest caz, mediana poate fi oricare din valorile cuprinse în intervalul (a_2, a_3).

Cvantilele corespunzătoare valorilor $k = 1/4$ și $k = 3/4$ se numesc respectiv *cvantila inferioară* și *cvantila superioară*

Scanned with OKEN Scanner

a distribuției. Ele se notează de obicei prin Q_1 (x) și Q_2 (x) și se construiesc grafic intersectînd curba $y = F(x)$ cu dreptele

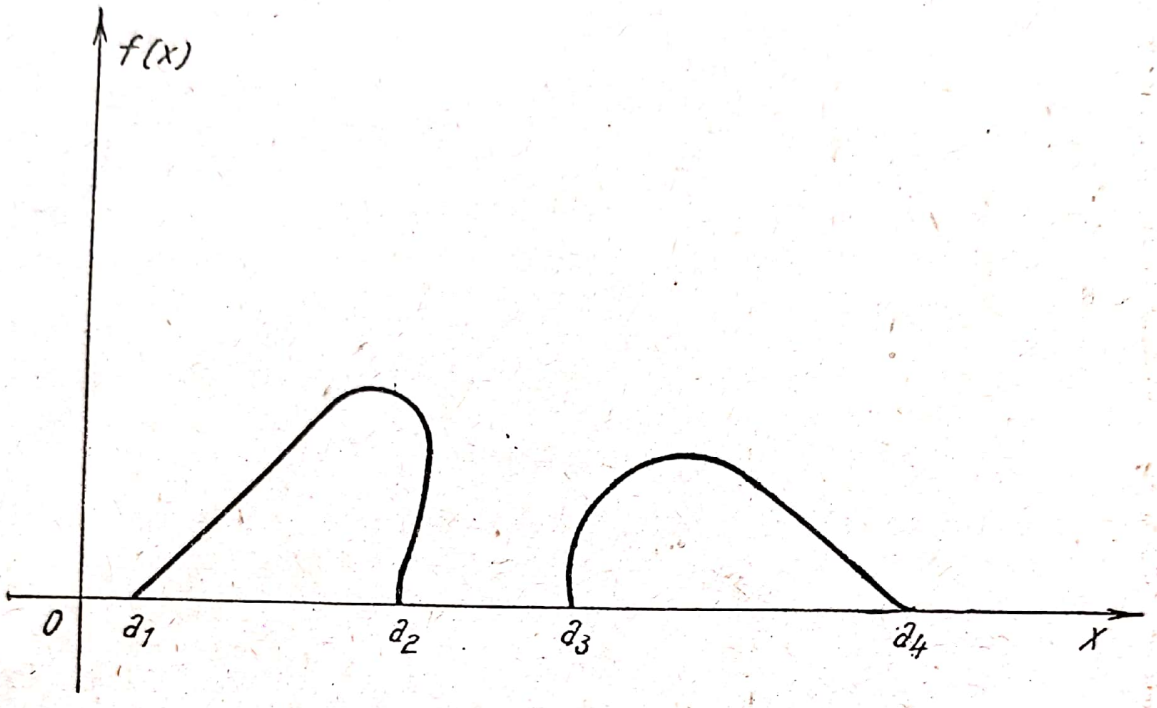

Fig. 35

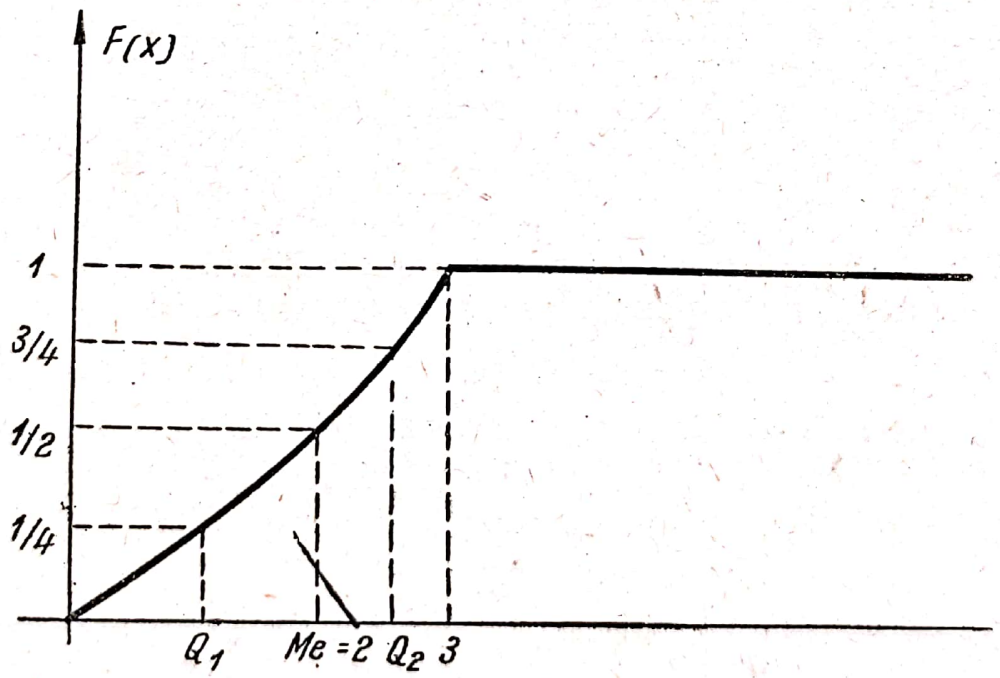

Fig. 36

$y = 1/4$ și $y = 3/4$. Dăm mai jos graficul cu mediana și cvantilele distribuției din exemplul de mai sus :

Scanned with OKEN Scanner

§ 2. Funcții monotone de variabile aleatorii

Variabila aleatorie x, prin relația :

$$u = u(x) \tag{6}$$

se transformă în variabila aleatorie u. Astfel o colectivitate în care variază x poate fi transformată în altă colectivitate în care variază u. Dacă într-o selecție din populația în care variază x s-a obținut valoarea x′ de pildă, aceasta se poate considera ca o observație u′ = u (x′) din populația u. *Densitatea de probabilitate a variabilei u (x) o vom nota g (u) și ea reprezintă probabilitatea ca variabila aleatorie u(x) să ia o valoare determinată u, adică* :

$$P\,[u(x) = u] = g(u) \tag{6_1}$$

Ne propunem să determinăm *densitatea g (u) a variabilei u(x), cunoscînd funcția u (x) și densitatea f(x) a variabilei x.* Ne limităm deocamdată la cazul particular cînd funcția u(x) este *monoton crescătoare.*

Dacă variabila u este de tip discret, presupunînd că ia valorile 0, 1, 2, ..., r cu probabilitățile respective f (0), f (1), f(2), ..., f(r), obținem imediat valorile u_0, u_1, u_2,,...u_r ale funcției u (x) substituind direct în (6) valorile lui x. Astfel $u_0 = u(0)$, $u_1 = u(1)$,...$u_r = u(r)$.

Deoarece funcția u (x) este monotonă, inversa sa

$$x = x^{-1}(u) \tag{6'}$$

este o funcție uniformă, deci probabilitatea ca u (x) să ia o anumită valoare u_k este tot una cu probabilitatea ca variabila x să ia valoarea corespunzătoare $x_k = x(u_k)$, deci :

$$g(u_k) = f(x_k) \tag{6_2}$$

Exemplu : Variabila x ia valorile 1, 2, 3, 4 cu probabilitățile f (1), f(2), f(3), f(4) și fie funcția monoton crescătoare

$$u(x) = (x-1)^2, \qquad x \geq 1$$

Valorile acestei funcții sînt 0, 1, 4, 9 cu probabilitățile respective date de relația (6_2) :

$$g(0) = f(1)$$
$$g(1) = f(2)$$
$$g(4) = f(3)$$
$$g(9) = f(4)$$

Scanned with OKEN Scanner

Să presupunem că variabila x din exemplul precedent este continuă, cu densitatea f(x). Graficul funcției $u(x) = (x-1)^2$ pentru $x \geqslant 1$ este cel de mai jos

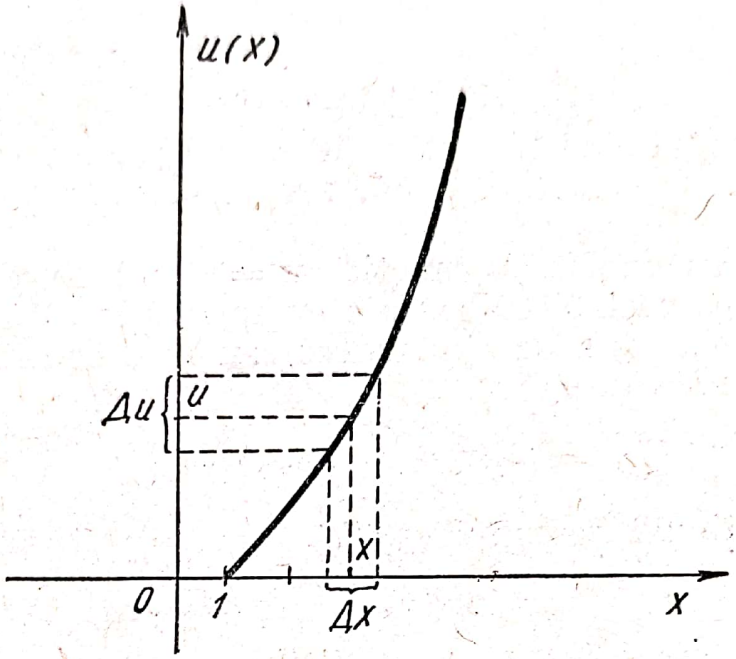

Fig. 37

Fie u o anumită valoare a funcției u (x) și Δu un interval oarecare în jurul punctului u. Valorii u îi corespunde valoarea x pe axa ox, iar intervalului Δu un interval Δx în jurul lui x. Avem evident $P(u \in \Delta u) = P(x \in \Delta x)$. Dar $P(u \in \Delta u)$ se obține după cum știm, integrînd funcția g (u) pe segmentul Δu;

$$P(u \in \Delta u) = \int_{\Delta u} g(u)\, du$$

și în mod analog,

$$P(x \in \Delta x) = \int_{\Delta x} f(x)\, dx$$

Aplicînd formula mediei obținem relația:

$$P(u \in \Delta u) = \int_{\Delta u} g(u)\, du = g(u')\, \Delta u$$

unde $u' \in \Delta u$, și în mod analog:

$$P(x \in \Delta x) = f(x')\, \Delta x,\ x' \in \Delta x$$

Avem deci relația: $g(u')\, \Delta u = f(x')\, \Delta x$

Scanned with OKEN Scanner

La limită, pentru $\Delta u \to 0$, $u' \to u$, $\Delta x \to 0, x' \to x$, deci putem scrie:

$$g(u) = f(x) \frac{dx}{du} \qquad (6'_2)$$

sau:

$$g(u)\, du = f(x)\, dx \qquad (6''_2)$$

Asupra cazului cînd funcţia $x(u)$ nu mai este uniformă, adică $u(x)$ nu este monotonă, vom reveni într-unul din capitolele următoare.

Valoarea medie a funcţiei $u = u(x)$ se calculează ca pentru orice variabilă aleatorie cu formulele (2) sau (4) după cum variabila este de tip discret sau este continuă.

Vom avea deci:

$$M(u) = \int_{-\infty}^{+\infty} u\, g(u)\, du$$

sau încă ţinînd seama de $(6''_2)$

$$M(u) = \int_{-\infty}^{+\infty} u(x)\, f(x)\, dx \qquad (6_3)$$

pentru variabila x continuă şi

$$M(u) = \sum u(x)\, f(x) \qquad (6'_3)$$

pentru variabila x, discretă; însumarea se face după toate valorile lui x.

Exemple. 1°. Fie $f(x) = C_3^x \left(\frac{1}{2}\right)^3$, $x = 0, 1, 2, 3$, şi $u(x) = x^2 + 1$.

Vom obţine valoarea medie (formula $6'_3$):

$$M(x^2 + 1) = \sum_{x=0}^{3} (x^2 + 1)\, C_3^x \left(\frac{1}{2}\right)^3 = 4$$

2°. Să considerăm funcţia $u(x) = (x - 1)^2$ în intervalul $(1, 4)$ şi densitatea lui x:

$$f(x) = \begin{cases} \frac{1}{12}(2x - 1), & 1 \leqslant x \leqslant 4 \\ 0, \text{ pentru celelalte valori} \end{cases}$$

Valoarea medie a funcţiei $u(x)$ se calculează cu formula (6_3):

$$M[(x-1)^2] = \int_1^4 (x-1)^2 \cdot \frac{1}{12}(2x - 1)\, dx$$

Scanned with OKEN Scanner

Pentru distribuțiile multidimensionale, valorile medii se definesc în mod analog ca și pentru cele unidimensionale. Fie, de exemplu, funcția :

$$u = u(x_1, x_2, \ldots, x_s) \tag{7}$$

unde distribuția s-dimensională cu variabilele $x_1, x_2, \ldots, x_s$ are densitatea $f(x_1, x_2, \ldots, x_s)$.

Valoarea medie a funcției u va fi dată de integrala multiplă

$$M(u) = \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} u(x_1, x_2, \ldots x_s) f(x_1, x_2, \ldots x_s) dx_1 \ldots dx_s \tag{7$_1$}$$

dacă variabilele $x_1, x_2, \ldots, x_s \ldots$ sînt continue, sau :

$$M(u) = \sum_{x_1} \ldots \sum_{x_s} u(x_1, x_2, \ldots x_s) f(x_1, \ldots x_s) \tag{7'$_1$}$$

dacă varțiabilele $x_1, x_2, \ldots x_s$ sînt de tip discret.

Sumele din partea dreaptă se extime asupra tuturor valorilor posibile ale variabilelor x_k $(1 \leqslant k \leqslant s)$.

Valorile medii ale distribuțiilor discrete sau continui au unele propietăți comune[1]. Adesea, pentru demonstrarea acestor proprietăți vom folosi distribuțiile continue.

C o r o l a r u l I. Dacă C este constanta, avem :

$$M(C) = C \tag{8}$$

În adevăr,

$$M(C) = \int_{-\infty}^{+\infty} Cf(x) dx = C \int_{-\infty}^{+\infty} f(x) dx = C. 1 = C$$

C o r o l a r u l II. Dacă u (x) și v (x) sînt funcții monotone de variabile aleatorii x

$$M[u(x) + v(x)] = M u(x)] + M [v(x)] \tag{8$_1$}$$

În adevăr după definiție avem :

$$M(u + v) = \int_{-\infty}^{+\infty} (u + v) f(x) dx = \int_{-\infty}^{+\infty} u f(x) dx + \int_{-\infty}^{+\infty} vf(x) dx =$$

$$= M(u) + M(v) \text{ (relația } 6_3)$$

În general, pentru s funcții de x, putem scrie

$$M\left\{\sum_{k=1}^{s} u_k(x)\right\} = \sum_{k=1}^{s} M[u_k(x)] \tag{8'$_1$}$$

T e o r e m a I. Valoarea medie a sumei de variabile aleatorii este egală cu suma valorilor medii ale fiecărei variabile, adică

[1] Justificarea acestor afirmații se poate obține folosind proprietățile integralei Stieljes.

$$M(x_1 + x_2 + \ldots + x_s) = M(x_1) + M(x_2 + \ldots + M(x_s) \qquad (8_2)$$

Demonstrație. Să considerăm mai întîi două variabile x_1 și x_2. După (7_1) vom avea :

$$E(x_1 + x_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x_1 + x_2) f(x_1, x_2)\, dx_2\, dx_1 =$$

$$= \int_{-\infty}^{+\infty} x_1 \left[\int_{-\infty}^{+\infty} f(x_1, x_2)\, dx_2 \right] dx_1 + \int_{-\infty}^{+\infty} x_2 \left[\int_{-\infty}^{+\infty} f(x_1, x_2)\, dx_1 \right] dx_2$$

Deoarece :

$$\int_{-\infty}^{+\infty} f(x_1, x_2)\, dx_2 = f_1(x_1), \int_{-\infty}^{+\infty} f(x_1, x_2)\, dx_1 = f_2(x_2)$$

sînt densitățile marginale în raport cu x_1, respectiv cu x_2, rezultă :

$$M(x_1 + x_2) = \int_{-\infty}^{+\infty} x_1 f_1(x_1)\, dx_1 + \int_{-\infty}^{+\infty} x_2 f_2(x_2)\, dx_2$$

Pentru distribuțiile unidimensionale x_1 și x_2, însă putem scrie (formula 4) :

$$\int_{-\infty}^{+\infty} x_1 f_1(x_1)\, dx_1 = M(x_1), \int_{-\infty}^{+\infty} x_2 f_2(x_2)\, dx_2 = M(x_2) \qquad (8_2')$$

Prin urmare,

$$M(x_1 + x_2) = M(x_1) + M(x_2) \qquad (8_2'')$$

Să demonstrăm acum (8_2) prin recurență. Admitem relația $M(x_1 + x_2 + \ldots + x_{s-1}) = M(x_1) + M(x_2) + \ldots + M(x_{s-1})$ și să demonstrăm că (8_2) este valabilă. Pentru aceasta vom scrie $x_1 + x_2 \ldots + x_{s-1} + x_s = x + x_s$ unde $x = x_1 + x_2 + \ldots + x_{s-1}$.

Atunci : $M(x_1 + x_2 + \ldots + x_s) = M(x + x_s) = M(x) + M(x_s)$ în baza relației $(8_2'')$. Dar :

$$M(x) = M(x_1 + x_2 + \ldots + x_{s-1}) = \sum_{k=1}^{s-1} M(x_k)$$

prin ipoteză. Înlocuind mai sus obținem relația (8_2).

T e o r e m a I I. Valoarea medie a unui produs de variabile aleatorii independente între ele este egală cu produsul valorilor medii ale fiecărei variabile :

$$M(x_1, x_2, \ldots, x_s) = M(x_1) M(x_2) \ldots M(x_s) \qquad (8_3)$$

Demonstrație. Prin definiție avem :

$$M(x_1 \dots x_s) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} x_1 \dots x_s f(x_1, \dots, x_s)\, dx_1 \dots dx_s$$

Deoarece variabilele sînt independente putem scrie după (17_2), capitolul V :

$$f(x_1 \dots, x_s) = \prod_{k=1}^{s} f_k(x_k)$$

Integrala multiplă se transformă atunci în produsul a s integrale simple, deci :

$$M(x_1 x_2 \dots x_s) = \int_{-\infty}^{+\infty} x_1 f_1(x_1)\, dx_1 \int_{-\infty}^{+\infty} x_2 f_2(x_2)\, dx_2 \dots \int_{-\infty}^{+\infty} x_s f(x_s)\, dx_s =$$

$$= M(x_1) \cdot M(x_2) \dots M(x_s)$$

deoarece :

$$\int_{-\infty}^{+\infty} x_k f_k(x_k)\, dx_k = M(x_k) \qquad (8'_3)$$

așa după cum am arătat mai înainte $(8'_2)$.

Observație. Independența variabilelor x_k este esențială pentru valabilitatea relației (8_3).

§ 3. Momentele distribuțiilor unidimensionale

D e f i n i ț i e. Valoarea medie a variabilei aleatorii $u(x) = x^r$ se numește *momentul de ordinul r al variabilei aleatorii x.*

În baza acestei definiții, notînd prin m_r acest moment, vom avea :

$$m_r = \int_{-\infty}^{+\infty} x^r f(x)\, dx \qquad (9)$$

sau :

$$m_r = \sum x^r f(x) \qquad (9')$$

dacă variabila x este de tip discret.

În particular, avem momentele de ordine 0 și 1 :

$$m_0 = \int_{-\infty}^{+\infty} f(x)\, dx = 1 \qquad (9_1)$$

$$m_1 = \int_{-\infty}^{+\infty} x f(x)\, dx = M(x) \qquad (9'_1)$$

Scanned with OKEN Scanner

Pentru uşurinţă vom nota adesea

$$m = m_1 = M(x) \qquad (9_1'')$$

Observaţie. Calculul momentelor definite mai sus se face socotind diferitele valori ale lui x de la originea 0. De aceea, li se mai spune şi *momente iniţiale.*

M o m e n t e c e n t r a t e. În statistica matematică este însă nevoie să calculăm momentele în raport şi cu alte puncte decît originea O. Îndeosebi este important să calculăm momentele unei distribuţii în raport cu valoarea sa medie $m = M(x)$.

D e f i n i ţ i e. Momentele calculate în raport cu valoarea medie a distribuţiei se numesc *momente centrate.*

Funcţia liniară $\xi(x) = x - m$ se numeşte *abatere.*

Din definiţia de mai sus, rezultă că *momentul centrat de ordinul r al variabilei x este momentul de ordinul r* (definit cu relaţia 9) al variabilei ξ. Notînd acest moment prin μ_r putem scrie :

$$\mu_r = M(\xi^r) = M[(x - m)^r] = \int_{-\infty}^{+\infty} (x - m)^r f(x)\, dx \qquad (9_3)$$

Pentru distribuţiile discrete avem :

$$\mu_r = \sum (x - m)^r f(x) \qquad (9_3')$$

Dacă se cunosc momentele iniţiale m_r, momentele centrate se calculează cu ajutorul acestora. În adevăr, putem scrie

$$\mu_r = M[(x - m)^r] = M\left[\sum_{k=0}^{r} (-1)^{r-k} C_r^k x^k m^{r-k}\right]$$

De unde, aplicînd teorema I :

$$\mu_r = \sum_{k=0}^{r} (-1)^{r-k} C_r^k m^{r-k} M(x^k)$$

sau încă ţinînd seama că $M(x^k) = m_k$, obţinem relaţia căutată

$$\mu_r = \sum_{k=0}^{r} (-1)^{r-k} C_r^k m^{r-k} m_k \qquad (9_4)$$

În particular, avem momentele centrate :

$$\mu_0 = m_0 = 1$$

$$\mu_1 = M(x - m) = M(x) - m = m - m = 0$$

$$\mu_2 = m_2 - m^2$$

Scanned with OKEN Scanner

Ultima relație se poate deduce din (9_4) făcînd $r = 2$:

$$\mu_2 = C_2^0\, m_0\, m^2 - C_2^1\, m_1\, m + C_2^2\, m_2 =$$

$$= m_2 - 2\, m^2 + m^2 = m_2 - m^2$$

sau pornind de la definiția momentului centrat :

$$\mu_2 = \int_{-\infty}^{+\infty} (x - m)^2\, f(x)\, dx = \int_{-\infty}^{+\infty} (x^2 - 2mx + m^2)\, f(x)\, dx =$$

$$= \int_{-\infty}^{+\infty} x^2\, f(x)\, dx - 2m \int_{-\infty}^{+\infty} x\, f(x)\, dx + m^2 \int_{-\infty}^{+\infty} f(x)\, dx =$$

$$= m_2 - 2\, m \cdot m + m^2 \cdot 1 = m_2 - m^2$$

D e f i n i ț i e. Momentul centrat de ordinul al doilea se numește *dispersia* sau *varianța* variabilei x. Se notează de obicei prin $D^2(x)$, deci :

$$D^2(x) = \mu_2 = \int_{-\infty}^{+\infty} (x - m)^2\, f(x)\, dx \qquad (9_5)$$

Rădăcina pătrată din dispersie se numește *abatere tip* sau *abatere medie pătratică* și se însemnează obișnuit prin σ :

$$\sigma = \sqrt{\mu_2} = D(x) \qquad (9_6)$$

C o r o l a r u l I. Dacă C este constanta

$$D^2(C) = 0 \qquad (9_7)$$

În adevăr,

$$D^2(C) = \int_{-\infty}^{+\infty} (C - m)^2\, f(x)\, dx \text{ . Dar } m = MC = C \text{ (relația 8)}$$

deci :

$$D^2(C) = \int_{-\infty}^{+\infty} 0 \cdot f(x)\, dx = 0$$

C o r o l a r u l I I. $D^2(Cx) = C^2\, D^2(x)$ $\qquad (9_8)$

În adevăr,

$$D^2(Cx) = \int_{-\infty}^{+\infty} (Cx - Cm)^2\, f(x)\, dx = C^2 \int_{-\infty}^{+\infty} (x - m)^2\, f(x)\, dx = C^2\, D^2(x)$$

T e o r e m a I I I. Dispersia sumei de variabile independente este egală cu suma dispersiilor variabilelor :

$$D^2(x_1 + x_2 + \ldots + x_s) = D^2(x_1) + D^2(x_2) \ldots + D^2(x_s) \qquad (9_9)$$

Scanned with OKEN Scanner

Demonstrație. Să notăm $M(x_1) = m_1, \ldots M(x_s) = m_s$. Prin definiție avem :

$$D^2(x_1 + \ldots + x_s) = \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} \left[\sum_{k=1}^{s} x_k - M\left(\sum_{k=1}^{s} x_k \right) \right]^2 f(x_1 \ldots x_s)\, dx_1 \ldots dx_s$$

$$= \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} \left[\sum_{k=1}^{s} (x_k - m_k) \right]^2 f(x_1, \ldots, x_s) dx_1 \ldots dx_s$$

deoarce :

$$\sum_{k=1} x_k - M\left(\sum_{k=1}^{s} x_k \right) = \sum_{k=1}^{s} \bar{x} - \sum_{k=1}^{s} M(x_k) =$$

$$= \sum_{k=1}^{s} x_k - \sum_{k=1}^{s} m_k = \sum_{k=1}^{s} (x_k - m_k)$$

Dezvoltînd pătratul de sub integrală obținem :

$$\left[\sum_{k=1}^{s} (x_k - m) \right]^2 = \sum_{k=1}^{s} (x_k - m_k)^2 + 2 \sum_{k=1} (x_k - m_k)(x_h - m_h)$$

De altă parte, variabilele x_k fiind independente putem scrie

$$f(x_1, \ldots, x_s) = \prod_{k=1}^{s} f_k(x_k)$$

Înlocuind în valoarea dispersiei $D^2(x_1, x_2 \ldots x_s)$ de mai sus se găsește

$$D^2(x_1 + x_2 + \ldots + x_s) = \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} \left[\sum_{k=1}^{s} (x_k - m_k)^2 \right] \prod_{k=1}^{s} f_k(x_k) \prod_{k=1}^{s} dx_k +$$

$$+ 2 \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} \left[\sum_{k,h} (x_k - m_k)(x_h - m_h) \right] \prod_{k=1}^{s} f_k(x_k) \prod_{k=1}^{s} dx_k =$$

$$= \sum_{k=1}^{s} \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} (x_k - m_k)^2 \prod_{k=1}^{s} f_k(x_k) \prod_{k=1}^{s} dx_k +$$

$$+ 2 \sum_{k,h} \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} (x_k - m_k)(x_h - m_h) \prod_{k=1}^{s} f_k(x_k) \prod_{k=1}^{s} dx_k$$

Scanned with OKEN Scanner

Suma a doua însă este nulă. In adevăr :

$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} (x_k - m_k)(x_h - m_h) \prod_{k=1}^{s} f_k(x_k) \prod_{k=1}^{s} dx_k =$$

$$= \int_{-\infty}^{+\infty} f_1(x_1)\, dx_1 \dots \int_{-\infty}^{+\infty} (x_k - m_k) f_k(x_k)\, dx_k \dots \int_{-\infty}^{+\infty} x_s f_s(x_s)\, dx_s$$

și cum integrala :

$$\int_{-\infty}^{+\infty} (x_k - m_k) f_k(x_k)\, dx_k = 0$$

pentru orice $k = 1, 2, \dots s$, rezultă că toți termenii sumei duble sînt nuli.

Pe de altă parte integrala

$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} (x_k - m_k)^2 \prod_{k=1}^{s} f_k(x_k) \prod_{k=1}^{s} dx_k =$$

$$= \int_{-\infty}^{+\infty} f_1(x_1)\, dx_1 \dots \int_{-\infty}^{+\infty} (x_k - m_k)^2 f_k(x_k)\, dx_k \dots \int_{-\infty}^{+\infty} f_s(x_s)\, dx_s =$$

$$= \int_{-\infty}^{+\infty} (x_k - m_k)^2 f_k(x_k)\, dx_k = D^2(x_k)$$

deoarece :

$$\int_{-\infty}^{+\infty} f_h(x_h)\, dx_h = 1 \text{ pentru } h = 1, 2, \dots, s$$

În concluzie, obținem rezultatul

$$D^2(x_1 + \dots . x_s) = \sum_{k=1}^{s} D^2(x_k)$$

adică relația (9_9).

După corolarul II și teorema de mai sus obținem relația mai generală

$$D^2\left(\sum_{k=1}^{s} a_k x_k\right) = \sum_{k=1}^{s} a_k^2 D^2(x_k) \qquad (9_{10})$$

În particular, pentru $a_1 = 1$, $a_2 = -1, \dots a_3 = \dots = a_p = 0$ găsim dispersia diferenței $x_1 - x_2$:

$$D^2(x_1 - x_2) = D^2(x_1) + D^2(x_2) \qquad (9_{11})$$

Scanned with OKEN Scanner

Dispersia caracterizează modul cum valorile variabilei x se grupează în jurul valorii medii m = M (x). Să considerăm în acest scop două distribuţii x_1 şi x_2 cu densităţile f (x_1) şi f (x_2) ambele avînd aceeaşi valoare medie m şi să presupunem că prima distribuţie este mai concentrată în jurul valorii medii decît a doua. Din punct de vedere geometric, aceasta înseamnă, că aria mărginită de curba f_1 (x) cel puţin într-un interval (a, b) în jurul mediei, este mai mare decît aria mărginită de curba f_2 (x).

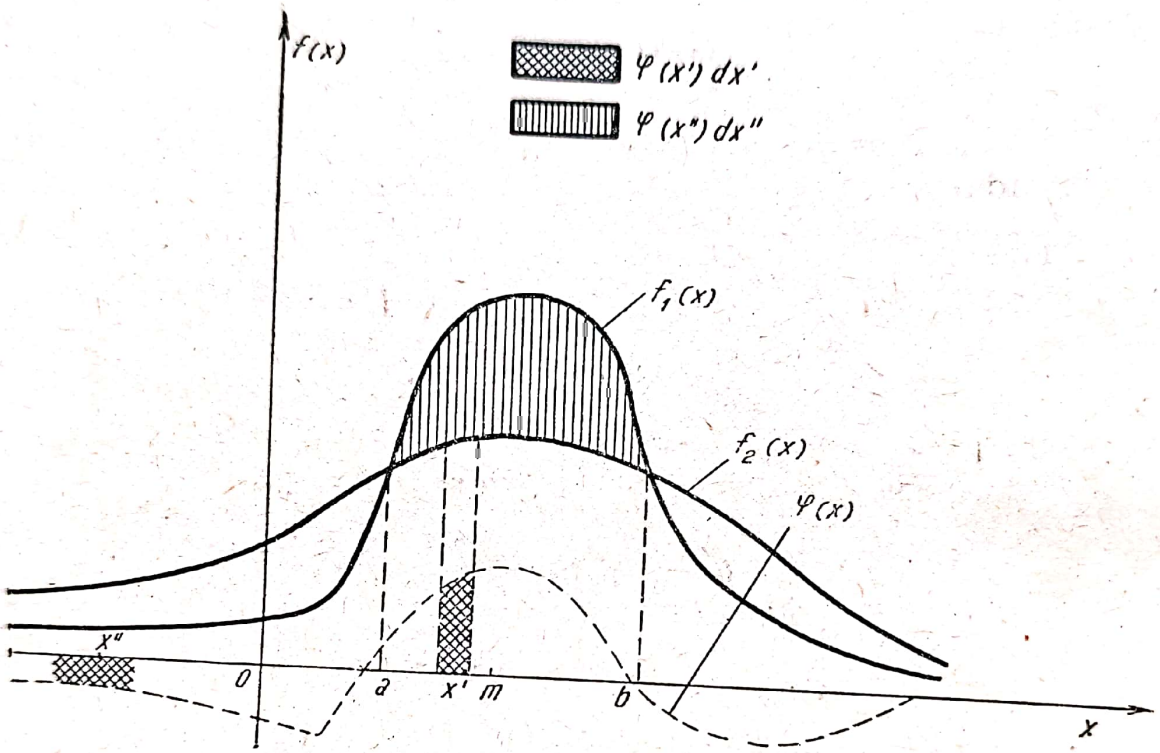

Fig. 38

Diferenţa acestor arii este haşurată în fig. 38. Această arie este egală cu integrala :

$$\int_b^a [f_1(x) - f_2(x)]\, dx > 0$$

Să considerăm acum curba a cărei ecuaţie este $\varphi(x) = f_1(x) - f_2(x)$. Deoarece avem :

$$\int_{-\infty}^{+\infty} \varphi(x)\, dx = \int_{-\infty}^{+\infty} f_1(x)(dx) - \int_{-\infty}^{+\infty} f_2(x)\, dx = 1 - 1 = 0$$

rezultă că ariile pozitive ale curbei $\varphi(x)$ cu axa ox se compensează cu cele negative, iar un element de arie pozitiv $\varphi(x')\, dx'$

Scanned with OKEN Scanner

se compensează cu un element de arie negativ $\varphi(x'')\,dx'''$ unde $|x'' - m| > |x' - m|$. Înmulțind elementele de arie respectiv prin $(x' - m)^2 < (x'' - m)^2$ și însumînd de la $-\infty$ la $+\infty$ obținem evident :

$$\int_{-\infty}^{+\infty} (x - m)^2 \varphi(x)\,dx < 0$$

sau ținînd seama că $\varphi(x) = f_1(x) - f_2(x)$:

$$\int_{-\infty}^{+\infty} (x - m)^2 f_1(x)\,dx < \int_{-\infty}^{+\infty} (x - m)^2 f_2(x)\,dx < 0$$

de unde rezultă $D^2(x_1) < D^2(x_2)$.

Așadar, *distribuția mai concentrată în jurul valorii medii are dispersia mai mică.*

D e f i n i ț i e. Variabila aleatorie

$$\eta(x) = \frac{\xi(x)}{\sigma} = \frac{x - m}{\sigma} \tag{10}$$

se numește *abatere normată.*

Momentele abaterii normate le vom nota ϑ_r, deci

$$\vartheta_r = M(\eta^r) = M\left[\left(\frac{x - m}{\sigma}\right)^r\right] = \int_{-\infty}^{+\infty} \left(\frac{x - m}{\sigma}\right)^r f(x)\,dx \tag{10$_1$}$$

Între momentele m_r, μ_r și ν_r există unele relații simple. Așa, de pildă, putem scrie :

$$\nu_r = \frac{1}{\sigma^r} \int_{-\infty}^{+\infty} (x - m)^r f(x)\,dx$$

de unde rezultă relația :

$$\mu_r = \sigma^r \cdot \nu_r \tag{10$_2$}$$

între momentele abaterii normate și cele centrate.

C o r o l a r. Abaterea normată are valoarea medie egală cu zero și dispersia egală cu 1.

$$M(\eta) = 0, \quad D^2(\eta) = 1 \tag{10$_3$}$$

În adevăr, din relația (10) putem scrie :

$$M(\eta) = \frac{1}{\sigma} M(x - m) = \frac{1}{\sigma} M(\xi) = 0$$

Scanned with OKEN Scanner

deoarece :

$$M(\xi) = \mu_1 = 0 \text{ relația } (9'_{11})$$

$$D^2(\eta) = D^2\left(\frac{x-m}{\sigma}\right) = \frac{1}{\sigma^2}[D^2(x) + D^2(m)] = \frac{\mu_2}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1$$

Definiție. O distribuție se numește *simetrică*, dacă pentru orice a subzistă relația

$$f(m-a) = f(m+a) \tag{11}$$

Aceasta înseamnă că ordonatele a două puncte așezate simetric față de dreapta $x = m$ sînt egale. În caz contrar, distribuția este *asimetrică*. De exemplu, distribuția normală este simetrică, deoarece :

$$f(m-a) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\frac{a^2}{\rho^2}} = f(m+a)$$

Teorema IV. Momentele centrate de ordinul impar ale unei distribuții simetrice sînt nule,

$$\mu_{2r+1} = 0 \tag{11$_1$}$$

Demonstrație. Avem prin definiție :

$$\mu_{2r+1} = \int_{-\infty}^{+\infty} (x-m)^{2r+1} f(x)\,dx = \int_{-\infty}^{m} (x-m)^{2r+1} f(x)\,dx +$$

$$+ \int_{m}^{+\infty} (x-m)^{2r+1} f(x)\,dx$$

În prima integrală să facem schimbarea :

$$x - m = -t \quad (t > 0), \quad dx = -dt$$

$$\int_{-\infty}^{m} (x-m)^{2r+1} f(x)\,dx = \int_{\infty}^{0} (-t)^{2r+1} f(m-t)\,(-dt) =$$

$$= \int_{0}^{\infty} (-1)^{2r+1} t^{2r+1} f(m-t)\,dt$$

În integrala a doua, deoarece $x > m$ vom lua $x - m = t\ (t > 0)$:

$$\int_{m}^{\infty} (x-m)^{2r+1} f(x)\,dx = \int_{0}^{\infty} t^{2r+1} f(m+t)\,dt$$

Scanned with OKEN Scanner

De unde rezultă :

$$\mu_{2r+1} = \int_0^{\infty} (-1)^{2r+1} t^{2r+1} f(m+t)\, dt + \int_0^{\infty} t^{2r+1} f(m+t)\, dt =$$

$$= \int_0^{\infty} [(-1)^{2r+1} + 1]\, t^{2r+1} f(m+t)\, dt = 0$$

deoarece $f(m-t) = f(m+t)$, întrucît distribuția este simetrică.

A s i m e t r i e. Pentru aprecierea asimetriei α a unei distribuții o vom compara cu o distribuție simetrică, de pildă, cu cea normală pentru care este firesc să luăm $\alpha = 0$. Mai mult, vom face ca ambele distribuții să aibă aceeași valoare medie și aceeași dispersie, cu alte cuvinte vom folosi abaterile lor normate.

Deoarece pentru distribuția normală $\mu_3 = 0$ se ia ca măsură a asimetriei momentul de ordinul al treilea al variabilei normate, adică :

$$\alpha_3 = \nu_3 = M\left[\frac{(x-m)^3}{\sigma}\right] = \int_{-\infty}^{+\infty} \left(\frac{(x-m)}{\sigma}\right)^3 f(x)\, dx$$

de unde formula asimetriei :

$$\alpha = \frac{\mu_3}{\sigma^3}. \qquad (12)$$

Dacă $\alpha < 0$ distribuția se spune că prezintă o *asimetrie la stînga*; invers dacă $\alpha > 0$ distribuția are o *asimetrie la dreapta* sau distribuția are o *asimetrie negativă*, respectiv *asimetrie pozitivă*

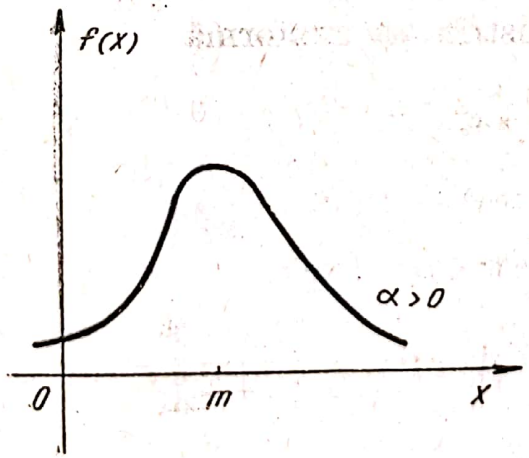

Fig. 39

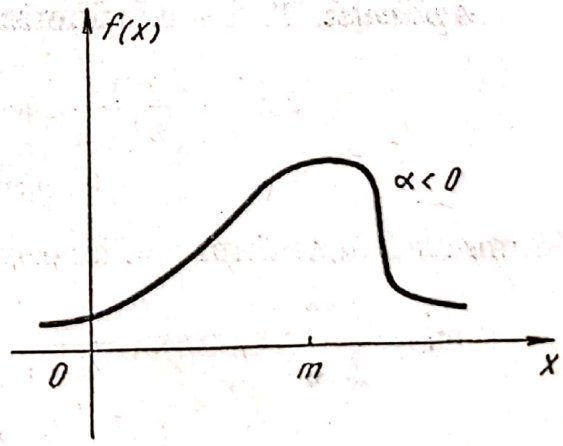

Fig. 40

Scanned with OKEN Scanner

Justificarea acestor denumiri este următoarea :

Dacă $\alpha = \frac{\mu_3}{\sigma 3} > 0$, cum $\sigma^3 > 0$ rezultă $\mu_3 > 0$; aceasta înseamnă că suma distanțelor $(x - m)^3$ pozitive este mai mare decît suma negativă, partea mai lungă a curbei se găsește la dreapta valorii medii și invers, dacă $\alpha < 0$.

Excesul sau *boltirea* β a unei distribuții este momentul de ordinul al patrulea al variabilei normate :

$$\beta = \nu_4 = \frac{\mu_4}{\sigma^4} \tag{13}$$

Pentru comparație se alege curba normală care are boltirea $\beta = 3$, după cum vom vedea mai departe. Dacă pentru o distribuție avem $\beta > 3$ curba are boltirea mai pronunțată decît a distribuției normale tip, iar dacă $\beta < 3$ curba este mai turtită :

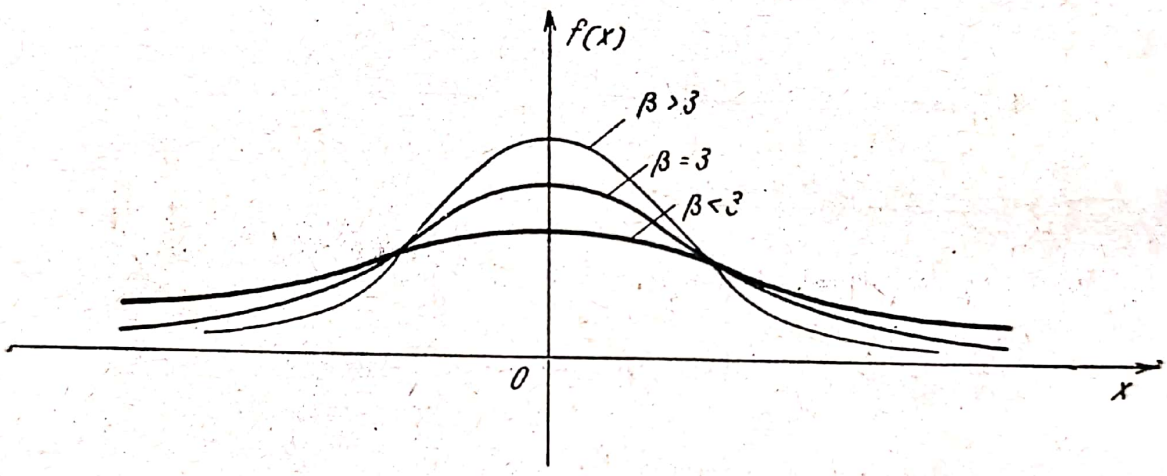

Fig. 41

Aplicații. 1°. Să considerăm distribuția uniformă

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{pentru } a \leqslant x \leqslant b \\ 0, & \text{pentru celelalte valori} \end{cases}$$

Momentul de ordinul r al acestei distribuții este :

$$m_r = \int_{-\infty}^{+\infty} x^r f(x)\, dx = \frac{1}{b-a} \int_a^b x^r\, dx = \frac{1}{b-a} \left[\frac{x^{r+1}}{r+1} \right]_a^b =$$

$$= \frac{b^{r+1} - a^{r+1}}{(r+1)(b-a)}$$

Scanned with OKEN Scanner

sau încă :

$$m_r = \frac{1}{r+1} \sum_{k=0}^{r} a^k b^{r-k}.$$

În particular, avem :

$$m_0 = 1,\ m_1 = \frac{1}{2}(a+b),\ m_2 = \frac{1}{3}(b^2 + ab + a^2)$$

De unde dispersia :

$$D^2(x) = \mu_2 = m_2 - m_1^2 = \frac{b^2+ab+a^2}{3} - \frac{a^2+b^2+2ab}{4} =$$

$$= \frac{a^2+b^2-2ab}{12} = \frac{1}{12}(b-a)^2$$

În acest caz, abaterea tip va fi :

$$\sigma = \sqrt{\mu_2} = D(x) = \frac{1}{2\sqrt{3}}(b-a)$$

Abaterea acestei distribuții este

$$\xi = x - m = x - \frac{a+b}{2}$$

iar abaterea normată :

$$\eta = \frac{\xi}{\sigma} = \left(x - \frac{a+b}{2}\right) : \frac{b-a}{2\sqrt{3}}$$

Distribuția este simetrică, deoarce : $f(m-c) = f(m+c) = \frac{1}{b-a}$ pentru orice punct c.

Momentele centrate se calculează cu formula

$$\mu_n = \int_{-\infty}^{+\infty} (x-m)^n f(x)\,dx = \int_a^b \left(x - \frac{a+b}{2}\right)^n \cdot \frac{1}{b-a}\,dx =$$

$$= \frac{1}{b-a}\left[\frac{\left(x - \frac{a+b}{2}\right)^{n+1}}{n+1}\right]_a^b = \frac{[1+(-1)^{n+1}](b-a)^n}{(n+1)\,2^{n+1}}.$$

Scanned with OKEN Scanner

Dacă $n = 2r$ avem :

$$\mu_{2r} = \frac{(b-a)^{2r}}{2^{2r}(2r+1)}.$$

În particular, pentru $r = 1$ avem $\mu_2 = \frac{(b-a)^2}{12}$ ca mai sus.

Dacă $n = 2r + 1$ rezultă $\mu_{2r+1} = 0$, după cum am constatat și în baza teoremei precedente. Momentele abaterii normate în acest caz, sînt :

$$\nu_{2r} = \frac{\mu_{2r}}{\sigma^{2r}} = \frac{3^r}{2r+1}$$

$$\nu_{2r+1} = 0$$

Dar asimetria distribuției este : $\alpha = \nu_3 = 0$ și boltirea ei :

$$\beta = \nu_4 = \frac{3^2}{2 \cdot 2 + 1} = \frac{9}{5} < 3$$

2°. Să considerăm distribuția normală tip

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \qquad -\infty < x < +\infty$$

Momentele sale vor fi :

$$\nu_n = \int_{-\infty}^{+\infty} x^n \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

Pentru calculul acestor momente, în integrala

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = 1$$

facem schimbarea :

$$x = y\sqrt{h}, \qquad dx = dy\sqrt{h}$$

și vom obține :

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = \frac{\sqrt{h}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{hy^2}{2}} dy = 1$$

sau :

$$\int_{-\infty}^{+\infty} e^{-\frac{hy^2}{2}} dy = \sqrt{2\pi}\, h^{-\frac{1}{2}}.$$

Scanned with OKEN Scanner

Derivînd de r ori în raport cu h se găseşte :

$$\frac{(-1)^r}{2^r}\int_{-\infty}^{+\infty} y^{2r}\, e^{-\frac{hy^2}{2}}\, dy = \sqrt{2\pi}\,(-1)^r \frac{1\cdot 3\cdot 5 \ldots (2r-1)}{2^r} h^{-\frac{r+2}{2}}.$$

Pentru $h = 1$ rezultă :

$$\nu_{2r} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} y^{2r} e^{-\frac{y^2}{2}}\, dy = 1\cdot 3\cdot 5 \ldots (2r-1)$$

Putem însă scrie înmulţind şi împărţind prin $2\cdot 4\cdot 6. \ldots (2r)$.

$$1\cdot 3\cdot 5\cdot \ldots 2r-1) = \frac{1\cdot 2\cdot 3\cdot 4\cdot 5 \ldots (2r-1)\,2r}{2\cdot 4\cdot 6 \ldots 2r} = \frac{(2r)!}{2^r\cdot r!}$$

aşa încît :

$$\nu_{2r} = \frac{(2r)!}{2^r\cdot r!} \qquad (14)$$

Am văzut mai înainte că $\nu_{2r+1} = 0$

Momentele centrate ale distribuţiei normale $N(x; m; \sigma)$ vor fi :

$$\mu_{2r} = \sigma^{2r}\nu_{2r} = \frac{(2r)!}{r!\,2^r}\cdot\sigma^{2r} \qquad (14')$$

$$\mu_{2r+1} = 0$$

Pentru $r = 1$ avem dispersia : $\mu_2 = D^2(x) = \sigma^2$

Deci parametrul σ din densitatea distribuţiei normale reprezintă abaterea tip a distribuţiei (în timp ce m reprezintă valoarea medie).

Boltirea distribuţiei normale se obţine făcînd $r = 2$, deci

$$\beta = \nu_4 = \frac{4!}{2^2\cdot 2!} = 3 \qquad (14_1)$$

după cum am afirmat mai înainte.

3°. Fie distribuţia binomială

$$f(x) = C_n^x\, p^x\, q^{n-x},\ x = 0,\ 1,\ 2,\ \ldots n$$

Valoarea sa medie $M(x) = np$ am calculat-o în paginile prece-

dente (relația 3). Momentul de ordinul r al acestei distribuții de tip discret se calculează după formula (2)

$$m_r = \sum_{x=0}^{n} x^r f(x) = \sum_{x=0}^{n} x^r C_n^x p^x q^{n-x}$$

Pentru calculul ultimei sume vom considera dezvoltarea

$$(pt + q)^n = \sum_{x=0}^{n} C_n^x p^x t^x q^{n-x}$$

pe care o derivăm de r ori în raport cu t și apoi facem $t = 1$. Obținem relația :

$$n(n-1)\ldots(n-r+1)\,p^r = \sum_{x=0}^{n} x(x-1)\ldots(x-r+1)\,C_n^x p^x q^{n-x}$$

Produsul de r factori de sub semnul sumă (Σ) se scrie :

$$P_r = \prod_{k=0}^{r-1} (x-k) = x^r - \sigma_r^1 x^{r-1} + \sigma_r^2 x^{r-2} + \ldots + (-1)^{r-1} \sigma_r^{r-1} x$$

Numerele σ_r^k $(1 \leqslant k \leqslant r-1)$ se numesc numerele lui Stirling de speța II și există tabele pentru valoarea lor. Dacă $r = 1$ avem $P_1 = x$, deci $\sigma_1^1 = 0$; pentru $r = 2$, $P_2 = x(x-1) = x^2 - x$, deci $\sigma_2^1 = 1$; pentru $r = 3$, $P_3 = x(x-1)(x-2) = x^3 - 3x^2 + 2x$ deci $\sigma_3^1 = 3$, $\sigma_3^2 = 2$, $\sigma_3^3 = 0$ etc.
Vom putea scrie relația :

$$n(n-1)\ldots(n-r+1)\,p^r = \sum_{x=0}^{n} x^r C_n^x p^x q^{n-x} - \sigma_r' \sum_{x=0}^{n} x^{r-1} C_n^x p^x q^{n-x} + \ldots$$

$$\ldots + (-1)^{r-1} \sigma_r^{r-1} \sum_{x=0}^{n} x\, C_n^x p^x q^{n-x}$$

sau încă :

$$n(n-1)\ldots(n-r+1)\,p^r = m_r - \sigma_r^1 m_{r-1} + \sigma_r^2 m_{r-2} + \ldots + (-1)^{r-1} \sigma_r^{r-1} m_1$$

Dacă facem $r = 1$ obținem valoarea medie $m = np = m_1$
Pentru $r = 2$ avem : $n(n-1)\,p^2 = m_2 - \sigma_2^1 m_1$
de unde rezultă ținînd seama că $\sigma_2^1 = 1$ și $1 - p = q$:
$m_2 = n^2 p^2 + npq$

Scanned with OKEN Scanner

Dispersia distribuției va fi : $D^2(x) = \mu_2 = m_2 - m^2 = n^2p^2 + npq - n^2p^2 = npq$ deci abaterea tip este : $\sigma = \sqrt{npq}$ (14_2)

Momentul de ordinul al treilea se deduce din relația :

$n(n-1)(n-2)p^3 = m_3 - 3m_2 + 2m_1 = m_3 - 3n^2p^2 - 3npq + 2npq$, de unde : $m_3 = p^3n^3 + 3n^2p^2q + 2np^3 + npq$

Momentele centrate vor fi, ținînd seama de (9_{11}) :

$$\mu_1 = 0$$

$$\mu_2 = npq = \sigma^2$$

$$\mu_3 = m_3 - 3m_2m_1 + 2m_1^2 = np^2 + np.$$

iar momentele abaterii normate :

$$\nu_1 = 0$$

$$\nu_2 = \frac{\mu_2}{\sigma^2} = 1$$

$$\nu_3 = \frac{\mu_3}{\sigma^3} = \frac{np(p+1)}{(npq)\,3/2} = \frac{p+1}{q^{3/2}\sqrt{np}}.$$

Altă metodă. Să notăm cu x_k numărul de cîte ori evenimentul A cu aceeași probabilitate p în fiecare probă, se realizează în proba de rangul k $(1 \leqslant k \leqslant n)$. Evident x_k este o variabilă aleatorie care poate lua valoarea 1 cînd s-a realizat A (probabilitatea respectivă p) și 0 cînd nu s-a realizat A (probabilitatea $q = 1 - p$). Numărul x de cîte ori se realizează A în n probe este evident egal cu suma

$$x = x_1 + x_2 + \ldots + x_n \qquad (14_3)$$

După teorema I avem :

$$M(x) = \sum_{k=1}^{n} M(x_k)$$

Dar $M(x_k) = 1.p + 0.q = p$, pentru orice k, deci :

$$M(x) = p + p + \ldots + p = np \text{ (formula 3)}$$

Momentul de ordinul al doilea va fi :

$$M(x^2) = M[(x_1 + x_2 + \ldots x_n)]^2 = M\left[\sum_{k=1}^{n} x_k^2 + 2\sum_{k,h} x_k x_h\right] =$$

$$= \sum_{k=1}^{n} M(x_k^2) + 2\sum_{k,h} M(x_k x_h).$$

Dar $M(x_k^2) = 1^2 \cdot p + 0^2 \cdot q = p$.

Scanned with OKEN Scanner

De altă parte variabilele x_k fiind independente după teorema II avem $M(x_k x_h) = M(x_k) \cdot M(x_h) = p^2$. Cum numărul termenilor sumei duble $\sum_{k,h} = \sum_{k} \sum_{h}$ este egal cu $C_n^2 = \frac{1}{2} n(n-1)$ vom obține: $m_2 = np + n(n-1)p^2 = n^2 p^2 + np(1-p) = n^2 p^2 + npq$ ca mai sus.

Dispersia distribuției se poate calcula cu ajutorul acestor momente cum am procedat mai înainte, însă se poate calcula și direct folosind teorema III $D^2(x) = D^2(x_1) + \ldots + D^2(x_n)$. Dar $D^2(x^k) = M(x_k - p)^2 = (1-p)^2 p + (0-p)^2 q = pq$ de unde rezultă $D^2(x) = \sigma^2 = npq$.

4°. *Distribuția hipergeometrică* are valoarea medie

$$M(x) = \frac{na}{m} = np \qquad \text{(formula } 3_2\text{)}$$

Momentul m_2 se calculează cu ajutorul formulei:

$$m_2 = \sum_{x=0}^{n} x^2 \frac{C_a^x C_{m-a}^{n-x}}{C_m^n} = \frac{\sum_{x=0}^{n} [x(x-1) + x] C_a^x C_{m-a}^{n-x}}{C_m^x} =$$

$$= \frac{1}{C_m^n} \left[\sum_{x=0}^{n} x(x-1) C_a^x C_{m-a}^{n-x} + \sum_{x=0}^{n} x C_a^x C_{m-a}^{n-x} \right]$$

A doua sumă din paranteză o avem calculată, valoarea ei este a C_{m-1}^{n-1}. Să calculăm prima sumă:

$$\sum_{x=0}^{n} x(x-1) C_a^x C_{m-a}^{n-x} = \sum_{x=0}^{n} x(x-1) \frac{a(a-1)\ldots(a-x+1)}{x!} C_{m-a}^{n-x} =$$

$$= \sum a(a-1) \frac{(a-2)\ldots(a-x+1)}{(x-2)!} C_{m-a}^{n-x} = a(a-1) \sum_{x=0}^{n} C_{a-2}^{x-2} C_{m-a}^{n-x}$$

În baza formulei (α) §1 avem însă:

$$\sum_{x=0}^{n} C_{a-2}^{x-2} \cdot C_{m-a}^{n-x} = C_{m-2}^{n-2}$$

De unde rezultă:

$$m_2 = \frac{a(a-1) C_{m-2}^{n-2} + a C_{m-1}^{n-1}}{C_m^n}$$

Scanned with OKEN Scanner

Efectuînd calculele se găseşte :

$$m_2 = \frac{a\,(a-1)\,n\,(n-1)}{m\,(m-1)} + \frac{an}{m}$$

De unde dispersia :

$$D^2(x) = \sigma^2 = m_2 - m_1^2 = \frac{a\,(a-1)\,n\,(n-1)}{m\,(m-1)} + \frac{an}{m} - \frac{a^2 n^2}{m^2} =$$

$$= n \cdot \frac{a}{m} \cdot \frac{m-a}{m} \cdot \frac{m-n}{m-1}$$

Notînd $\frac{a}{m} = p$ obţinem :

$$\sigma^2 = npq\,\frac{m-n}{m-1} \qquad (14_4)$$

5. *Corecţiile lui Shepard.* În primul capitol am arătat că momentelor empirice li se aduc anumite corecţii numite corecţiile lui Shepard.

În cele ce urmează, vom arăta cum se calculează aceste corecţii.

Mai înainte însă vom reaminti o formulă cunoscută sub numele de formula de însumare a lui Euler — Mac-Laurin :

Fie $\varphi(x)$ o funcţie continuă pe intervalul (a, b) ; să împărţim acest interval în n părţi egale prin punctele

$$a,\ a+h,\ a+2h,\ \ldots,\ a+(n-1)h,\ a+nh = b.$$

Avem relaţia :

$$h\sum_a^b \varphi(x_i) = \int_a^b \varphi(x)\,dx + \sum_{i=1}^{\infty} A_i \left| \varphi_{(x)}^{(i-1)} \right|_a^b \qquad (14_5)$$

unde am notat :

$$\sum_a^b \varphi(x_i) = \varphi(a) + \varphi(a+h) + \varphi(a+2h) + \ldots + \varphi(b)$$

$$\left| \varphi_{(x)}^{(i-1)} \right|_a^b = \varphi_{(b)}^{(i-1)} - \varphi_{(a)}^{(i-1)}$$

$\varphi_{(x)}^{(i-1)}$ fiind derivata de ordinul $(i-1)$ a funcţiei $\varphi(x)$, iar $A_1, A_2, \ldots$ nişte coeficienţi constanţi.

Scanned with OKEN Scanner

După cum se știe, momentul de ordinul k al variabilei aleatorii x de tip continuu, este dat de formula :

$$m_k = \int_{-\infty}^{+\infty} x^k f(x)\,dx \qquad (14_6)$$

iar noi, în realitate putem calcula cu datele selecției momentul empiric :

$$m'_k = \sum_{h=1}^{r} \alpha_h^k f(x_h) \qquad (14_7)$$

unde am însemnat prin $f(x_h)$ frecvența valorilor lui x cuprinse în intervalul $\left(\alpha_h - \frac{d}{2},\ \alpha_h + \frac{d}{2}\right)$. Valoarea exactă a acestei frecvențe este evident

$$f(\alpha_h) = \int_{\alpha_h - \frac{d}{2}}^{\alpha_h + \frac{d}{2}} f(x)\,dx$$

unde f (x) este densitatea variabilei x.

Dacă facem schimbarea $x = t + \alpha_h$, $dx = dt$ integrala de mai sus devine :

$$f(\alpha_h) = \int_{-\frac{d}{2}}^{+\frac{d}{2}} f(\alpha_h + t)\,dt$$

sau încă, dezvoltînd în serie Taylor funcția $f(\alpha_h + t)$ și integrînd obținem :

$$f(\alpha_h) = \int_{-\frac{d}{2}}^{+\frac{d}{2}} \left\{ f(\alpha_h) + tf'(\alpha_h) + \frac{t^2}{2} f''(\alpha_h) + \ldots + \frac{t^m}{m!} f^{(m)}(\alpha_h) + \right\} dt$$

$$= \left[t f(\alpha_h) + \frac{t^2}{2} f^1(\alpha_h) + \frac{t^3}{3!} f''(\alpha_h) + \ldots + \frac{t^{m+1}}{(m+1)!} f^{(m)}_{(\alpha_h)} + \ldots \right]_{-\frac{d}{2}}^{\frac{d}{2}}.$$

Deoarece pentru m impar avem :

$$\left[t^{m+1} \right]_{-\frac{d}{2}}^{\frac{d}{2}} = 0$$

Scanned with OKEN Scanner

vom putea scrie frecvența f (α_h) sub forma :

$$f(\alpha_h) = \sum_{q=0}^{\infty} \frac{d^{2q+1}}{(2q+1)!\, 2^{2q}} f^{(2q)}_{(\alpha_h)}$$

Momentul selecției (14_7) devine atunci :

$$m'_k = \sum_{h=1}^{r} \alpha_h^k \sum_{q=0}^{\infty} \frac{d^{2q+1}}{(2q+1)!\, 2^{2q}} f^{(2q)}_{(\alpha_h)} =$$

$$= \sum_{q=0}^{\infty} \frac{d^{2q+1}}{(2q+1)!\, 2^{2q}} \sum_{h=1}^{r} \alpha_h^k f^{(2q)}_{(\alpha_h)} \qquad (14_8)$$

Vom folosi acum formula Euler — Mac Laurin în care înlocuim pe a cu a_1 și b cu a_{r+1}, valoarea cea mai mică, respectiv cea mai mare a scării intervalelor (vezi relația (4), capitolul I). Făcînd în relația (14_5) :

$$\varphi(x_i) = \alpha_i^k f^{(2q)}_{(\alpha_i)} ; \; h = d$$

obținem relația :

$$d \sum_{h=1}^{r} \alpha_h^k f^{(2q)}_{(\alpha_h)} = \int_{a_1}^{a_{r+1}} x^k f^{(2q)}(x)\, dx + \sum_{i=0}^{\infty} A_i \left| x^k f^{(i-1)}(x) \right|_{a_1}^{a_{r+1}}$$

Întrucît în afara intervalului (a_1, a_{r+1}) nu există frecvențe relative, putem neglija seria din partea dreaptă și obținem :

$$\sum_{h=1}^{r} \alpha_h^k f^{(2q)}_{(\alpha_k)} = \frac{1}{d} \int_{a_1}^{a_{r+1}} x^k f^{(2q)}_{(x)}\, dx$$

Momentul m'_k devine :

$$m'_k = \sum_{q=0}^{\infty} \frac{d^2 q}{(2q+1)!\, 2^{2q}} \int_{a_1}^{a_{r+1}} x^k f^{(2q)}_{(x)}\, dx$$

Notînd $x^k = u$, $f^{(2q)}_{(x)}\, dx = dv$, integrala devine :

$$\int_{a_1}^{a_{r+1}} x^k f^{(2q)}_{(x)}\, dx = \left| x^k f^{(2q-1)}_{(x)} \right|_{a_1}^{a_{r+1}} - k \int_{a_1}^{a_{r+1}} x^{k-1} f^{(2q-1)}_{(x)}\, dx =$$

$$= -k(k-1) \int_{a_1}^{a_{r+1}} x^{k-r} f^{(k-r)}_{(x)}\, dx =$$

$$= k(k-1) \ldots (k-2q+1) \int_{a_1}^{a_{r+1}} x^{k-2q} f(x)\, dx$$

Scanned with OKEN Scanner

dacă $2q \leqslant k$ și

$$\int_{a_1}^{a_{r+1}} x^k f_{(x)}^{(2q)} dx = 0$$

dacă $2q > k$.

În baza acestor rezultate, formula (14_8) se poate scrie:

$$m'_k = \sum_{2q \leqslant k} \frac{k(k-1)\ldots(k-2q+1)}{(2q+1)!} \left(\frac{d}{2}\right)^{2q} m_{k-2q} \qquad (14_9)$$

În particular, vom avea:

$$m'_0 = m_0, \quad m'_1 = m_1, \quad m'_2 = m_2 + \frac{d^2}{12} m_0 \qquad (14_{10})$$

$$m'_3 = m_3 + \frac{d^3}{4} m_1, \quad m'_4 = m_4 + \frac{d^2}{2} m_2 + \frac{d^4}{80} m_0$$

iar pentru momentele centrate:

$$\mu_2 = \mu'_2 - \frac{d^2}{12}, \quad \mu_3 = \mu'_3, \quad \mu_4 = \mu'_4 - \frac{d^2}{2} \mu'_2 + \frac{7d^4}{240} \qquad (14_{11})$$

§ 4. Funcția generatoare de momente

Pentru uniformitatea calculului momentelor și pentru unele studii teoretice în legătură cu distribuțiile statistice se introduce noțiunea de funcție generatoare de momente pe care o vom nota prin $m(t)$.

Prin definiție, *valoarea medie a variabilei aleatorii* e^{tx}, unde t este un număr real, *se numește funcția generatoare a momentelor variabilei aleatorii* x. Avem deci prin definiție:

$$m(t) = M(e^{tx}) = \int_{-\infty}^{+\infty} e^{tx} f(x)\, dx \qquad (15)$$

unde $f(x)$ este densitatea variabilei x.

Funcția generatoare se bucură de următoarele proprietăți:

1°. $m(0) = 1$ (15_1)

2°. Dacă $x = x_1 + x_2 + \ldots + x_n$ și x_k sînt independente între ele, avem:

$$M(e^{tx}) = \prod_{k=1}^{n} M(e^{tx_k})$$

Scanned with OKEN Scanner

sau încă :

$$m(t) = m_1(t) \cdot m_2(t) \ldots m_n(t) \qquad (15_2)$$

unde $m_k(t) = M(e^{txk})$ = funcţia generatoare a variabilei x_k $(1 \leqslant k \leqslant n)$.

În adevăr, după teorema II, putem scrie :

$$M(e^{t(x_1 + x_2 + \ldots + x_n)} = M(e^{tx_1} e^{tx_2} \ldots e^{tx_n}) =$$
$$= M(e^{tx_1}) \cdot M(e^{tx_2}) \ldots M(e^{tx_n}) \qquad (15_3)$$

3°. $m(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} m_k$

dacă variabila x admite momente finite de orice ordin.
Avem din (15), dezvoltînd e^{tx} în serie :

$$m(t) = M\left(1 + \frac{t}{1!} x + \frac{t^2}{2!} x^2 + \ldots + \frac{t^k}{k!} x^k + \ldots\right) =$$

$$= \sum_{x=0}^{\infty} \frac{t^k}{k!} M(x^k) = \sum_{x=0}^{\infty} \frac{t^k}{k!} m_k \qquad (15_4)$$

$$4°. \quad = m_{(o)}^{(k)}, m_k; \; k = 1, 2, 3, \ldots$$

unde $m^{(k)}(0)$ este derivata de ordinul k în raport cu t, a funcţiei m (t) în care facem apoi t = 0 :

$$m^{(k)}(0) = [m^{(k)}(t)]_{t=0} \qquad (15_5)$$

În adevăr, dezvoltînd m (t) în serie Mac-Laurin şi ţinînd seama de (15_3) obţinem relaţia :

$$\sum_{x=0}^{\infty} \frac{t^k}{k!} m^{(k)}(0) = \sum_{x=0}^{\infty} \frac{t^k}{k!} m_k$$

de unde prin identificare rezultă (15_4).

Aşadar, *cunoscînd funcţia generatoare, prin derivări succesive în raport cu t şi făcînd apoi t = 0 se obţin toate momentele distribuţiei*, în ipoteza bineînţeles, că ele există (sînt finite).

Exemple. 1°. Pentru distribuţia lui Poisson avem funcţia generatoare :

$$m(t) = M(e^{tx}) = \sum_{x=0}^{\infty} e^{tx} f(x) = \sum_{x=0}^{\infty} e^{tx} e^{-a} \frac{a^x}{x!} =$$

$$= e^{-a} \sum_{x=0}^{\infty} \frac{(a e^t)^x}{x!} = e^{-a} \cdot e^{a e^t}$$

deci :

$$m(t) = e^{a(e^t - 1)} \tag{16}$$

De aici deducem derivatele :

$$m'(t) = a\,e^t\,e^{a(e^t-1)}$$

$$m''(t) = a\,e^t \cdot e^{a(e^t-1)} + a^2 e^{2t}\, e^{a(e^t-1)}$$

Făcînd $t = 0$ avem momentele :

$$m_1 = a,\ m_2 = a + a^2 \tag{16$_1$}$$

de unde dispersia :

$$D^2_{(x)} = m_2 - m_1^2 = a + a^2 - a^2 = a \tag{16$_2$}$$

2°. Distribuția binomială are funcția generatoare :

$$m(t) = M(e^{tx}) = \sum_{x=0}^{n} e^{tx} \cdot C_n^x\, p^x\, q^{n-x} = \sum_{x=0}^{n} C_n^x\,(e^t p)^x\, q^{n-x}$$

Dar

$$\sum_{x=0}^{n} C_n^x\,(p\,e^t)^x\, q^{n-x} = (p\,e^t + q)^n\,,$$

deci :

$$m(t) = (p\,e^t + q)^n \tag{16$_3$}$$

Avem derivatele :

$$m'(t) = n\,p\,(p\,e^t + q)^{n-1}\,e^t$$

$$m''(t) = n(n-1)\,p^2\,(p\,e^t + q)^{n-2}\,e^{2t} + np\,(p\,e^t + q)^{n-1}\,e^t$$

De unde momentele :

$$m_1 = m'(0) = np$$

$$m_2 = n(n-1)\,p^2 + np = n^2 p^2 + npq$$

și dispersia : $\sigma^2 = m_2 - m_1^2 = npq$

3°. Pentru distribuția uniformă avem funcția :

$$m(t) = \int_b^a e^{tx}\,\frac{1}{b-a}\,dx$$

De unde deducem :

$$m^{(r)}\,t = \frac{1}{b-a}\int_a^b x^r\,e^{tx}\,dx$$

Scanned with OKEN Scanner

deci :

$$m_r = m^{(r)}(0) = \frac{1}{b-a}\int_a^b x^r\,dx = \frac{1}{b-a}\left(\frac{b^{r+1}-a^{r+1}}{r+1}\right)$$

Pentru $r = 1$, se găsește $m_1 = \frac{1}{2}(a+b)$.

4°. *Funcția generatoare pentru distribuția Gama* este :

$$m(t) = \int_0^\infty e^{tx}\,\frac{x^a\,e^{-\frac{x}{b}}}{\Gamma(a+1)\cdot b^{a+1}}\,dx = \frac{1}{(1-bt)^{a+1}} \qquad (16_5)$$

În adevăr, putem scrie :

$$m(t) = \frac{\int_0^\infty x^a\,e^{-x\left(\frac{1-bt}{b}\right)}\,dx}{\Gamma(a+1)b^{a+1}} = \frac{\Gamma(a+1)\,b^{a+1}}{(1-bt)^{a+1}}\cdot\frac{1}{\Gamma(a+1)\,b^{a+1}} =$$

$$= \frac{1}{(1-bt)^{a+1}}$$

Derivata de ordinul r în raport cu t, va fi : $m^{(r)}(t) = (a+1)(a+2)\ldots(a+r)(1-bt)^{-(a+r+1)}\,b^r$ de unde obținem momentul căutat : $m^{(r)}(0) = m_r = (a+1)(a+2)\ldots(a+r)\,b^r$

Funcția generatoare a momentelor centrate. Funcția generatoare a abaterii $\xi = x - m$ este :

$$m_1(t) = M[e^{t\xi}] = M[e^{t(x-m)}] = e^{-mt}\,M(e^{tx})$$

sau :

$$m_1(t) = e^{-mt}\,m(t) \qquad (17)$$

Deci, *înmulțind funcția generatoare a variabilei x prin factorul* e^{-mt}, *obținem funcția generatoare a momentelor centrate.*

După proprietatea 4° a funcției generatoare, avem :

$$\mu_r = m_1^{(r)}(0) \quad \text{pentru } r = 1, 2, \ldots \qquad (17_1)$$

Exemplu : Pentru distribuția binomială am găsit : $m = np$, $m(t) = (pe^t + q)^n$
deci :

$$m_1(t) = e^{-npt}(pe^t + q)^n = [e^{-pt}(pe^t + q)]^n = (pe^{qt} + qe^{-pt})^n$$

Avem derivatele :

$$m_1'(t) = n\,(pe^{qt} + qe^{-pt})^{n-1}\,(pqe^{qt} - pqe^{-pt})$$

$$m_1''(t) = n\,(n-1)\,(pe^{qt} + qe^{-pt})^{n-2}\,(pqe^{qt} - pqe^{-pt})^2 +$$

$$+ n\,(pe^{qt} + qe^{-pt})^{n-1}\,(pq^2e^{qt} + p^2qe^{-pt})$$

De unde deducem momentele centrate :

$$\mu_1 = m_1'(0) = 0$$

$$\mu_2 = \sigma^2 = m_1''(0) = npq.$$

Funcția generatoare pentru momentele, abaterii normate $\eta = \frac{x-m}{\sigma}$.

Notînd prin $m_2(t)$ această funcție avem :

$$m_2(t) = M(e^{t\eta}) = M\left[e^{t\frac{(x-m)}{\sigma}}\right] = e^{-m\frac{t}{\rho}}\,M\left(e^{\frac{t}{\sigma}x}\right)$$

sau încă :

$$m_2(t) = e^{-\frac{mt}{\sigma}}\,m\left(\frac{t}{\sigma}\right) \qquad (18)$$

adică înlocuim în $m(t)$ pe t prin $\frac{t}{\sigma}$ și înmulțim cu factorul $e^{-\frac{mt}{\sigma}}$. Momentele variabilei normate vor fi :

$$\nu_r = m_2^{(r)}(0) \qquad (18_1)$$

Aceste momente se folosesc îndeosebi cînd densitatea distribuției este exprimată în funcție de m și σ.

Exemplu : Să considerăm distribuția normală. Funcția sa generatoare este

$$m(t) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{tx}\cdot e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}\,dx = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{tx-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}\,dx .$$

Să facem schimbarea : $\frac{x-m}{\sigma} = u$, $dx = \sigma du$, atunci avem :

$$m(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{t(m+\sigma u)-\frac{1}{2}u^2}\,du$$

Scanned with OKEN Scanner

Exponentul lui e se poate scrie :

$$\frac{2\,tm + 2\,t\,\sigma\,u - u^2}{2} = -\frac{u^2 - 2\,t\,\sigma\,u + t^2\sigma^2 - t^2\sigma^2 - 2\,tm}{2} =$$

$$= -\frac{(u - t\sigma)^2 - (2\,m\,t + t^2\sigma^2)}{2}$$

rezultă atunci :

$$m\,(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{mt + \frac{t^2\sigma^2}{2}}\, e^{-\left(\frac{u-t\sigma}{2}\right)^2} du =$$

$$= \frac{e^{mt + \frac{t^2\sigma^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(u-t\sigma)^2} du$$

Întrucît integrala :

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(u-t\sigma)^2} du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}z^2} dz = 1$$

obținem funcția generatoare :

$$m\,(t) = e^{mt + \frac{t^2\sigma^2}{2}} \qquad (18_2)$$

Funcția generatoare a abaterii normate va fi :

$$m_2\,(t) = e^{-m\frac{t}{\sigma}} : e^{m\frac{t}{\sigma} + \frac{t^2\sigma^2}{2\sigma^2}} = e^{\frac{t^2}{2}} \qquad (18_3)$$

Această funcție se putea de altfel obține direct din (18_2) făcînd $m = 0$, $\sigma = 1$.

Dezvoltînd $m_2\,(t)$ cu formula lui Mac-Laurin avem :

$$m_2(t) = M\,(e)^{t\eta} = \sum_{x=0}^{\infty} \frac{t^{2k}}{(2\,k)\,!}\, \nu_{2k}$$

Pe de altă parte putem scrie dezvoltarea :

$$e^{\frac{t^2}{2}} = 1 + \frac{t^2}{2 \cdot 1\,!} + \frac{(t^2)^2}{2^2 \cdot 2\,!} + \cdots + \frac{(t^2)^k}{2^k \cdot k\,!} + \cdots$$

Prin identificare obținem relația :

$$\frac{\nu\,2\,k}{(2\,k)\,!} = \frac{1}{2^k \cdot k\,!}$$

Scanned with OKEN Scanner

De unde :

$$\nu_{2k} = \frac{(2k)!}{k!\,2^k}, \quad \nu_{2k+1} = 0$$

Momentele centrate vor fi :

$$\mu_{2k} = \frac{(2k)!}{k!\,2^k}\,\sigma^{2k}, \quad \mu_{2k+1} = 0$$

adică formulele găsite mai înainte pe altă cale (14 și 14′). Momentele m_r se deduc din ν_r în mod simplu. În adevăr, din relația $\eta = \frac{x - m}{\sigma}$ deducem : $x = m + \sigma\eta$.

De unde :

$$m_r = M(x^r) = M\left[(m + \sigma\eta)^r\right] = M\left[\sum_{x=0}^{r} C_r^x\, m^x\, \sigma^{r-x}\, \eta^{r-x}\right]$$

sau :

$$m_r = \sum_{x=0}^{r} C_r^x\, m^x\, \sigma^{r-x}\, \nu_{r-x} \qquad (18_4)$$

§ 5. Momentele distribuțiilor multidimensionale

Noțiunile introduse în paragrafele precedente se extind și la distribuțiile multidimensionale.

Aici vom avea, pe lîngă momentele fiecărei variabile în parte și *momente mixte*.

Fie de pildă o distribuție de două variabile continue x și y, cu densitatea de probabilitate f (x, y). Momentul de ordinul k al distribuției, în raport cu variabila x se definește prin relația :

$$m_{k0} = M(x^k) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^k\, f(x, y)\, dx\, dy \qquad (19)$$

În mod analog, momentul de ordinul h în raport cu y este :

$$M(y^h) = m_{0h} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y^h\, f(x, y)\, dx\, dy \qquad (19')$$

Momentul mixt de ordinul k în raport cu x și de ordinul h în raport cu y, adică momentul de ordinul k+h al distribuției va fi

$$M(x^k y^h) = m_{kh} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^k\, y^h\, f(x, y)\, dx\, dy \qquad (19'')$$

Scanned with OKEN Scanner

În particular, avem valoarea medie în raport cu x :

$$m_{10} = M(x) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x\, f(x,y)\, dx\, dy = \int_{-\infty}^{+\infty} x \left[\int_{-\infty}^{+\infty} f(x, y)\, dy \right] dx =$$

$$= \int_{-\infty}^{+\infty} x\, f(x)\, dx \qquad (19_1)$$

unde f (x), reprezintă după cum se știe, densitatea marginală în raport cu x. În mod analog avem :

$$m_{01} = M(y) = \int_{-\infty}^{+\infty} y\varphi(y)\, dy \qquad (19'_1)$$

unde φ(y) este densitatea marginală în raport cu y. În același mod se definesc și momentele centrate. Așa, de pildă, momentul centrat de ordinul k, în raport cu x, va fi :

$$\mu_{k0} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - m_{10})^k f(x, y)\, dx\, dy \qquad (20)$$

iar momentul centrat de ordinul h, în raport cu y, este :

$$\mu_{0h} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - m_{01})^h f(x, y)\, dx\, dy \qquad (20')$$

Momentul centrat mixt, de ordinul k în raport cu x și de ordinul h în raport cu y va fi de asemenea :

$$\mu_{kh} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - m_{10})^k (y - m_{01})^h f(x, y)\, dx\, dy \qquad (20'')$$

În particular, avem :

$$\mu_{10} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - m_{10}) f(x, y)\, dx\, dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x\, f(x, y)\, dx\, dy -$$

$$- m_{10} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)\, dx\, dy = m_{10} - m_{10} = 0$$

după (19_1). Avem de asemenea $\mu_{01} = 0$.

Prin definiție, momentul centrat mixt de ordinul al doilea

$$\mu_{11} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - m_{10})(y - m_{01}) f(x, y)\, dx\, dy \qquad (20_1)$$

se numește *covarianța variabilelor* x și y.

Scanned with OKEN Scanner

Covarianța se mai poate încă scrie :

$$\mu_{11} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x\,y\,f(x,y)\,dx\,dy - m_{10}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y\,f(x,y)\,dx\,dy -$$

$$- m_{01}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x\,f(x,y)\,dx\,dy + m_{10}\,m_{01}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,dx\,dy =$$

$$= m_{11} - m_{10}\cdot m_{01} - m_{01}\cdot m_{10} + m_{10}\cdot m_{01}$$

De unde :

$$\mu_{11} = m_{11} - m_{10}\cdot m_{01} \qquad (20_2)$$

Dacă variabilele x și y sînt independente atunci : $f(x, y) = f(x)\,\varphi(y)$, iar covarianța se scrie :

$$\mu_{11} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x - m_{10})(y - m_{01})\,f(x)\,\varphi(y)\,dx\,dy =$$

$$= \int_{-\infty}^{+\infty} (x - m_{10})\,f(x)\,dx \int_{-\infty}^{+\infty} (y - m_{01})\,\varphi(y)\,dy = 0$$

Așadar, *dacă variabilele sînt independente covarianța lor este nulă.*

Dacă în formulele (20) și (20′) facem $k = h = 2$, obținem dispersiile :

$$\mu_{20} = \mu_x = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x - m_{10})^2\,f(x,y)\,dx\,dy \qquad (21)$$

$$\mu_{02} = \mu_y = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (y - m_{01})^2\,f(x,y)\,dx\,dy \qquad (21')$$

Dezvoltînd parantezele se găsesc relațiile :

$$\mu_{20} = m_{20} - m_{10}^2 \cdot \quad \mu_{02} = m_{02} - m_{01}^2 \qquad (21_1)$$

asemănătoare cu relația $\mu_2 = m_2 - m_1^2$ din cazul unidimensional.

Corelația variabilelor x și y se definește ca raportul

$$\rho = \frac{\mu_{11}}{\sqrt{\mu_{20}\cdot\mu_{02}}} \qquad (21_2)$$

Abaterile tip ale variabilelor x și y vor fi :

$$\sigma_x = \sqrt{\mu_{20}}, \quad \sigma_y = \sqrt{\mu_{02}} \qquad (21_3)$$

Scanned with OKEN Scanner

Cu aceste notații corelația ρ se mai poate scrie :

$$\rho = \frac{\mu_{11}}{\sigma_x \sigma_y} \qquad (21'_2)$$

Numărului ρ, definit mai sus, i se spune și *coeficient de corelație.* El verifică relația $-1 \leqslant \rho \leqslant +1$.
În adevăr, se consideră forma pătratică pozitiv definită

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [\alpha (x - m_{10}) + \beta (y - m_{01})]^2 f(x, y)\, dx\, dy =$$

$$= \mu_{20} \alpha^2 + 2 \mu_{11} \alpha\beta + \mu_{02} \beta$$

Pentru ca trinomul din partea doua să fie pozitiv este necesar ca $\mu_{11}^2 - \mu_{20} \mu_{02} < 0$, de unde deducem :

$$\rho^2 = \frac{\mu_{11}^2}{\mu_{20} \mu_{02}} < 1$$

adică obținem relația care trebuia demonstrată.

Funcția generatoare de momente pentru o distribuție bidimensională se definește în modul următor :

$$m(t_1, t_2) = M(e^{t_1 x + t_2 y}) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{t_1 x + t_2 y} f(x, y)\, dx\, dy \qquad (22)$$

Momentul de ordinul k în raport cu x se va obține derivînd funcția $m(t_1, t_2)$ de k ori în raport cu t_1 și făcînd apoi $t_1 = t_2 = 0$, adică :

$$m_{k0} = \left[\frac{\partial^k}{\partial t_1^k} m(t_1, t_2)\right]_{t_1 = t_2 = 0} = \frac{\partial^k}{\partial t_1^k} m(0, o) \qquad (22_1)$$

În mod analog momentul de ordinul h în raport cu y se va obține derivînd $m(t_1, t_2)$ de h ori, în raport cu t_2 și făcînd apoi $t_1 = t_2 = 0$:

$$m_{0h} = \left[\frac{\partial^h}{\partial t_2^h} m(t_1, t_2)\right]_{t_1 = t_2 = 0} = \frac{\partial^h}{\partial t_2^h} m(0, o) \qquad (22'_1)$$

Momentul mixt m_{kh} se obține derivînd $m(t_1, t_2)$ de k ori în raport cu t_1, de h ori în raport cu t_2 și făcînd apoi $t_1 = t_2 = 0$:

$$m_{kh} = \frac{\partial^{k+h}}{\partial t_1^k \partial t_2^h} m(0, o) \qquad (22''_1)$$

Fie acum o distribuție cu s variabile $x_1, x_2 \ldots x_s$ și cu densitatea de probabilitate $f(x_1, x_2, \ldots, x_s)$.

Momentul de ordinul k în raport cu x_r va fi :

$$M\left(x_r^k\right) = \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} x_r^k \, f(x_1, \ldots x_s)\, dx_1 \ldots dx_s \qquad 1 \leqslant r \leqslant s \tag{23}$$

iar momentul mixt de ordinul k, în raport cu x_r, de ordinul h în raport cu x_u și de ordinul l în raport cu x_v, va fi

$$M\left(x_r^k x_u^h x_v^l\right) = \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} x_r^k x_u^h x_v^l \, f(x_1, \ldots x_s)\, dx_1 \ldots dx_s \qquad 1 \leqslant r, u, v \leqslant s \tag{23'}$$

Funcția generatoare de momente a distribuției se definește ca în relația :

$$m(t_1, t_2, \ldots t_s) = M\left(e^{t_1 x_1 + \ldots + t_s x_s}\right) =$$

$$= \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} e^{\sum_{k=1}^{s} t_k x_k} f(x_1, \ldots x_s)\, dx_1 \ldots dx_s \tag{23''}$$

Diferitele momente ale distribuției se calculează prin derivări succesive. Avem, de pildă :

$$M\left(x_r^k\right) = \frac{\partial^k}{\partial t_r^k} m(0, 0, \ldots 0) \tag{23_1}$$

$$M\left(x_r^k x_u^h x_v^l\right) = \frac{\partial^{k+h+1}}{\partial t_r^k \partial_u^h \partial t_v^l} m(0, 0, \ldots 0) \tag{$23_1'$}$$

Covariațele și corelațiile celor s variabile luate două cîte două se definesc ca mai sus.

§ 6. Metoda funcției caracteristice

Calculul momentelor unei distribuții statistice se poate face tot așa de simplu dacă în locul funcției generatoare se folosește funcția caracteristică, definită ca valoarea probabilă a funcției e^{itx}, unde i este simbolul imaginar, iar t un parametru real. În baza acestei definiții, notînd prin $\varphi(t)$ funcția caracteristică a variabilei x, avem :

$$\varphi(t) = M\left(e^{itx}\right) = \int_{-\infty}^{+\infty} e^{itx} f(x)\, dx \tag{24}$$

Scanned with OKEN Scanner

sau :

$$\varphi(t) = \sum e^{itx} f(x) \qquad (24')$$

dacă variabila x este de tip discret; însumarea se face după toate valorile posibile ale variabilei → x.

Funcția caracteristică are un cîmp mult mai mare de aplicare decît funcția generatoare, deoarce produsul $e^{tx} f(x)$ în multe cazuri nu tinde către zero cînd x tinde către infinit.

Totuși, pentru distribuțiile cele mai des întîlnite în aplicațiile de care ne ocupăm în cursul de față, se poate folosi și funcția generatoare, așa încît, vom întrebuința atît funcția caracteristică cît și funcția generatoare, după împrejurări.

Ca și funcția generatoare, funcția caracteristică se bucură de anumite proprietăți dintre care enunțăm pe cele mai importante :

1°. $\varphi(0) = 1$ $\qquad (24_1)$

În adevăr, din (24) făcînd $t = 0$ găsim :

$$\varphi(0) = \int_{-\infty}^{+\infty} f(x)\,dx = 1.$$

2°. Dacă $x_1, x_2, \ldots, x_s$ sînt variabile aleatorii independente cu funcțiile caracteristice $\varphi_1(t), \varphi_2(t) \ldots \varphi_s(t)$, funcția caracteristică a sumei $x_1 + x_2 + \ldots + x_s$ este :

$$\varphi(t) = \varphi_1(t).\ \varphi_2(t) \ldots \varphi_s(t) \qquad (24_2)$$

În adevăr, după definiție avem

$$\varphi(t) = M\left[e^{it(x_1 + \ldots + x_s)}\right] = M\, e^{itx_1} \ldots e^{itx_s}] =$$

$$M(e^{itx_1}).\ M(e^{itx_2}) \ldots M(e^{itx_s})$$

deoarece variabilele sînt independente (teorema II). Pentru orice $k = 1, 2, \ldots, s$ putem însă scrie : $\varphi_k(t) = M(e^{tix_k})$ deci înlocuind în relația precedentă obținem relația (24_2)

3°. Este uniform continuă pentru orice t real.

Să formăm diferența $\varphi(t + h) - \varphi(t)$. Vom avea :

$$\varphi(t+h) - \varphi(t) = \int_{-\infty}^{+\infty} e^{i(t+h)x} f(x)\,dx - \int_{-\infty}^{+\infty} e^{itx} f(x)\,dx =$$

$$- \int_{-\infty}^{-a} e^{i(t+h)x} f(x)\,dx + \int_{a}^{\infty} e^{i(t+h)x} f(x)\,dx -$$

$$- \int_{-8}^{-a} e^{itx} f(x)\,dx - \int_{a}^{\infty} e^{itx} f(x)\,dx + \int_{-a}^{a} e^{itx}(e^{ihx} - 1) f(x)\,dx$$

Dacă luăm a și h în așa mod ca :

$$\int_{-\infty}^{-a} f(x)\,dx < \varepsilon/5,\quad \int_{a}^{+\infty} f(x)\,dx < \varepsilon/5,\quad \left| e^{iha} - 1 \right| < \varepsilon/5$$

vom obține : $|\varphi(t+h) - \varphi(t)| < \varepsilon$ oricare ar fi t.

4°. Avem dezvoltările :

$$\varphi(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!}\,\varphi^{(k)}(0) = \sum_{k=0}^{\infty} \frac{(ti)^k}{k!}\,m_k \qquad (24_3)$$

În adevăr, ținînd seama că $|e^{itx}| = 1$ obținem $|\varphi(t)| \leqslant 1$, deci $\varphi(t)$ este analitică în cercul de rază 1 cu centrul în origine și, ca atare, obținem dezvoltarea în serie Mac-Laurin :

$$\varphi(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!}\,\varphi^{(k)}(0)$$

Pe de altă parte avem :

$$M(e^{itx}) = M\left(\sum_{k=0}^{\infty} \frac{(it)^k}{k!}\,x^k\right) = \sum_{k=0}^{\infty} \frac{i^k t^k}{k!}\,M(x^k) = \sum_{x=0}^{\infty} \frac{i^k t^k}{k!}\,m^k$$

5. Momentele m_k ale distribuției verifică relațiile

$$m_k = \frac{1}{i_k}\left[\varphi^{(k)}(t)\right]_{t=0} = \frac{1}{i^k}\,\varphi^{(k)}(0)\quad k = 0,\ 1,\ 2,\ \ldots \qquad (24_4)$$

Această relație rezultă din (24_3) prin identificare.
Exemple. 1°. Pentru distribuția lui Poisson avem :

$$\varphi(t) = M(e^{itx}) = \sum_{x=0}^{\infty} e^{itx}\,e^{-m}\,\frac{m^x}{x!} =$$

$$= \sum_{x=0}^{\infty} e^{-m}\,\frac{(me^{it})^x}{x!}\,e^{-m} \sum_{x=0}^{\infty} \frac{(me^{it})^x}{x!} = e^{-m}\,e^{me^{it}}$$

sau :

$$\varphi(t) = e^{m(e^{it}-1)} \qquad (24_5)$$

Valoarea medie a acestei distribuții este, după relația (24_4)

$$m_1 = \frac{1}{i}\,\varphi'(0)$$

Scanned with OKEN Scanner

Deoarece $\varphi'(t) = mi\, e^{it} e^{m(e^{it}-1)}$, se găsește $m_1 = \frac{1}{i} m\, i = m$.

În mod analog se calculează și celelalte momente.

2°. Pentru distribuția Gama avem :

$$\varphi(t) = \int_0^{\infty} e^{itx} \frac{x^a e^{-\frac{x}{b}}}{\Gamma_{(a+1)}\, b^{a+1}}\, dx$$

Derivînd de k ori, în raport cu t, se găsește succesiv :

$$\varphi^{(k)}(t) = \int_0^{\infty} i^k\, e^{itx} \frac{x^{a+k} e^{-\frac{x}{b}}}{\Gamma_{(a+1)}\, b^{a+1}}\, dx$$

$$\varphi^{(k)}(0) = \frac{i^k}{\Gamma(a+1)\, b^{a+1}} \int_0^{\infty} x^{a+k} e^{-\frac{x}{b}}\, dx$$

$$= \frac{i^k}{\Gamma(a+1)\, b^{a+1}} \cdot \frac{\Gamma(a+k+1)}{\frac{1}{b^{a+k+1}}} =$$

$$= \frac{i^k\, b^{a+k+1} (a+k)(a+k-1) \ldots (a+1)\, \Gamma(a+1)}{\Gamma(a+1)\, b^{a+1}} =$$

$$= i^k (a+1)(a+2) \ldots (a+k)\, b^k$$

De unde momentul : $m_k = \frac{1}{i^k} \varphi^{(k)}(0) = (a+1)(a+2) \ldots$ $\ldots (a+k)\, b^k$ adică valoarea găsită mai înainte prin funcția generatoare.

Funcția caracteristică pentru momentele centrate va fi :

$$\varphi_1(t) = M\left[e^{it\xi}\right] = M\left[e^{it(x-m)}\right] = e^{-itm}\, \varphi(t) \tag{25}$$

Deci, în acest caz, vom obține :

$$\mu_k = \frac{1}{i^k} \varphi_1^{(k)}(0) \tag{25$_1$}$$

În mod analog, pentru variabila normată avem funcția caracteristică :

$$\varphi_2(t) = M\left[e^{it\eta}\right] = M\left[e^{it\left(\frac{x-m}{\sigma}\right)}\right] = e^{-i\frac{t}{\sigma}m}\, \varphi\left(\frac{t}{\sigma}\right) \tag{26}$$

iar pentru momentele acestei variabile

$$\nu_k = \frac{1}{i!}\varphi_2^{(k)}(0) \qquad (26_1)$$

Funcția caracteristică pentru o distribuție cu s variabile $x_1, x_2, \ldots x_s$, avînd densitatea de probabilitate $f(x_1, x_2 \ldots x_s)$ se definește ca și funcția generatoare :

$$\varphi(t_1 t_2, \ldots t_s) = M\left[e^{it_1 x_1 + it_2 x_2 + \ldots + it_s x_s}\right] =$$

$$= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} e^{i\sum_{k=1}^{s} st_k x_k} f(x_1, \ldots x_s)\, dx_1 \ldots dx_s \qquad (27)$$

Momentul de ordinul k în raport cu x_1, de pildă, va fi :

$$M(x_1^k) = \frac{1}{i^k}\frac{\partial^k}{\partial t_1^k}\varphi^{(k)}(0,0, \ldots 0) \qquad (27_1)$$

iar momentul mixt de ordinul k în raport cu x_1 și de ordinul h în raport cu x_2 este :

$$M(x_1^k x_2^h) = \frac{1}{i^{k+h}} \cdot \frac{\partial^{k+h}}{\partial t_1^k \partial t_2^h}\varphi^{(k+h)}(0, 0, \ldots, 0) \qquad (27_2)$$

§ 7. Problema momentelor

Din dezvoltările precedente s-a văzut că o distribuție statistică conduce la șirul de momente $m_0, m_1 \ldots$ Cu ajutorul acestora obținem funcția generatoare sau funcția caracteristică a distribuției și reciproc din acestea prin derivări succesive obținem momentele distribuției.

Una din problemele fundamentale ale statisticii matematice constă în *determinarea densității* $f(x)$ *cînd se cunoaște funcția generatoare de momente* $m(t)$ *sau funcția caracteristică* $\varphi(t)$, sau încă în *determinarea densității* $f(x)$ *prin cunoașterea momentelor distribuției.*

Prima demonstrație riguroasă a acestei probleme a fost dată de către Cebîșev și simplificată de Marcov. Soluția problemei a fost dată pe cale algebrică sub forma unei metode de apro-

ximații succesive. În esență, problema pusă și rezolvată de Cebîșev este următoarea :

Să se determine mulțimea de numere reale x_1, x_2,... și mulțimea numerelor pozitive $f(x_1)$, $f(x_2)$... asociate acestora în așa fel ca să satisfacă relațiile :

$$\begin{aligned} f(x_1) + f(x_2) + \ldots &= m_0 \\ x_1 f(x_1) + x_2 f(x_2) + \ldots &= m_1 \\ x_1^2 f(x_1) + x_2^2 f(x_2) + \ldots &= m_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots & \end{aligned} \tag{28}$$

unde m_0, m_1 m_2... sînt numere reale date.

Marea însemnătate a funcției caracteristice constă în aceea că ea rezolvă simplu și complet problema pusă la început.

T e o r e m a I. Din relația :

$$\varphi(t) = \int_{-\infty}^{+\infty} e^{itx} f(x)\, dx$$

aplicînd formula de inversiune a lui Fourier găsim :

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx} \varphi(t)\, dt \tag{29}$$

Vom demonstra această teoremă arătînd că relația (29) se reduce la o identitate cînd înlocuim $\varphi(t)$ prin valoarea sa de mai sus.

În acest scop calculăm integrala $\int_{-a}^{+a} e^{-itx} \varphi(t)\, dt$

în care facem apoi $a \to \infty$. Vom avea :

$$\int_{-a}^{+a} e^{-itx} \varphi(t)\, dt = \int_{-a}^{a} e^{-itx} dt \int_{-\infty}^{+\infty} f(u)\, e^{itu} du =$$

$$= \int_{-a}^{a} dt \int_{-\infty}^{+\infty} e^{i(u-x)t} f(u)\, du = \int_{-a}^{a} dt \int_{-\infty}^{+\infty} [\cos(u-x)t + i \sin(u-x)t\, f(u) du$$

Dar integrala :

$$\int_{-\infty}^{+\infty} \sin(u-x)\, t\, f(u)\, du = 0$$

întrucît funcția sin (u—x) t este o funcție impară, deci :

$$\int_{-a}^{a} e^{-itx} \varphi(t)\, dt = \int_{-a}^{a} dt \int_{-\infty}^{+\infty} \cos(u-x)\, t f(u)\, du =$$

$$= \int_{-\infty}^{+\infty} f(u)\, du \int_{-a}^{a} \cos(u-x)\, t\, dt = 2 \int_{-\infty}^{+\infty} \frac{\sin a(u-x)}{u-x} f(u)\, du$$

deoarece :

$$\int_{-a}^{a} \cos(u-x)\, t\, dt = \left[\frac{\sin(u-x)\, t}{u-x} \right]_{-a}^{a} = \frac{\sin a(u-x)}{u-x} - \frac{\sin[-a(u-x)]}{u-x}$$

$$= \frac{\sin a(u-x)}{u-x} + \frac{\sin a(u-x)}{u-x} = 2 \frac{\sin a(u-x)}{u-x}$$

Făcînd substituția $a(u-x) = \theta$, $du = \frac{1}{a} d\theta$ găsim :

$$\int_{-a}^{a} e^{-itx} \varphi(t)\, dt = 2 \int_{-\infty}^{+\infty} \frac{\sin \theta}{\theta} f\left(\frac{\theta}{a} + x \right) d\theta$$

Pentru $a \longrightarrow \infty$ integrala din partea dreaptă tinde către limita $2\pi f(x)$, adică :

$$\lim_{\substack{-a \to a \\ a \to \infty}} \int_{-a}^{a} e^{-itx} \varphi(t)\, dt = 2\pi f(x) = \int_{-\infty}^{+\infty} e^{-itx} \varphi(t)\, dt$$

de unde se găsește pentru f (x) valoarea (29).

Funcția de repartiție se obține, după cum știm, integrînd funcția (29) între limitele $-\infty$ și x.

Așadar, cunoașterea funcției caracteristice este echivalentă cu cunoașterea distribuției însăși.

Teorema II. Se dau două distribuții continue cu densitățile f (x) și φ (x).

Dacă momentele celor două distribuții sînt egale și dacă diferența $\delta(x) = f(x) - \varphi(x)$ este analitică în jurul originei, cele două densități sînt echivalente.

Demonstrație. Prin ipoteză avem : $\delta(x) = a_0 + a_1 x + a_2 x^2 \ldots$ În condițiile teoremei, integrala :

$$\int_{-\infty}^{+\infty} [\delta(x)]^2\, dx = 0$$

Scanned with OKEN Scanner

În adevăr, putem scrie :

$$\int_{-\infty}^{+\infty} \left[\delta(x)\right]^2 dx = \int_{-\infty}^{+\infty} \delta(x)\,(a_0 + a_1 x + a_2 x^2 + \ldots)\,dx =$$

$$= a_0 \int_{-\infty}^{+\infty} [f(x) - \varphi(x)]\,dx + a_1 \int_{-\infty}^{+\infty} x\,[f(x) - \varphi(x)]\,dx + \ldots .$$

$$= a_0 (1 - 1) + a_1 (m_1 - m_1) + a_2 (m_2 - m_2) + \ldots = 0$$

Deoarece : $[\delta(x)]^2 \geqslant 0$ și cum integrala este nulă rezultă : $\delta(x) = f(x) - \varphi(x) = 0$

C o r o l a r. Dacă două variabile au aceeași funcție generatoare de momente sau aceeași funcție caracteristică, ele au aceeași densitate de probabilitate.

În adevăr, dacă variabilele au aceeași funcție generatoare momentele lor coincid după cum rezultă din relația :

$$m(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} m_k = \sum_{k=0}^{\infty} \frac{t^k}{k!} m'_k$$

unde m_k și m'_k sînt momentele de ordinul k ale celor două variabile.

Probleme

1. Să se reprezinte grafic densitatea Gama pentru :
1°. $a = 4,\ b = 1$
2°. $a = 2,\ b = 1$
Pentru ambele cazuri să se calculeze μ_3/σ^3

2. Să se calculeze raportul μ_4/σ_4 pentru distribuția Beta și să se arate că această distribuție este simetrică în punctul $x = \frac{1}{2}$ dacă $a = b$.

3. Prin definiție avem :

$$\int_{-\infty}^{+\infty} \frac{1}{\pi} \frac{x\,dx}{1 + (x - m)^2} = \lim_{\lambda \to \infty} \int_{m-\lambda}^{m+\lambda} \frac{1}{\pi} \frac{x\,dx}{1 + (x - \lambda)^2}$$

Să se arate că distribuția lui Cauchy nu are momente de ordin superior.

4. Se dă variabila aleatorie x cu densitatea

$$f(x) = 2\sqrt{\frac{2}{\pi}}\, x\, e^{-\frac{x^4}{2}},\ x > 0$$

Să se calculeze valoarea medie și dispersia variabilei x.

Scanned with OKEN Scanner

5. Variabila x are momentele :

$$m_r = r\,! : \left(\frac{r}{2}\right)!\text{, dacă r este par ;}$$
$$= 0\text{, dacă r este impar}$$

Să se determine densitatea variabilei x din funcția ei generatoare.

6. Dacă lnx are distribuția normală cu media 1 și dispersia 4, să se calculeze $P\ (\frac{1}{2} < x < 2)$.

7. Momentele definite de funcția generatoare $w(t) = \ln m(t)$ se numesc *invarianții distribuției*. Notînd aceste momente prin $s_1, s_2 \ldots s_r \ldots$ avem $s_r = w^{(r)}(0)$, derivata fiind calculată în raport cu t. Să se demonstreze realațiile : $s_1 = m_1, s_2 = \sigma^2$.

8. Să se calculeze invariantul de ordinul r, s_r pentru densitatea $f(x) = k\ e^{kx}$, $x > 0$.

9. Să se demonstreze că invarianții $s_1', s_2' \ldots s_r' \ldots$ ai funcției $y = ax$ sînt dați de relațiile $s_r' = a^r s_r$, $r = 1, 2 \ldots$ unde s_r sînt invarianții variabilei x.

10. Se numește *moment factorial de ordinul s al variabilei x* valoarea probabilă a produsului $x(x-1)\ldots(x-s+1)$. Notîndu-l prin $m_{(s)}$ avem prin definiție $m_{(s)} = M[x(x-1)\ldots(x-s+1)]$. Funcția $w(t) = M(t^x)$ se numește *funcție generatoare de momente factoriale* ale variabilei x.

Dacă x este bornată și ia numai valori întregi $x = 0,1,2 \ldots m$, notînd $P(x = k) = p_k$ vom avea $w(t) = \sum_{k=0}^{m} p_k t^k$, sau încă dezvoltînd în serie Taylor :

$$w(t) = w(1) + \sum_{s=1}^{m} \frac{w^{(s)}(1)}{s\,!}(t-1)^s$$

Cum $w(1) = 1$ se găsește :

$$w^{(s)}(1) = \sum_{k=s}^{m} p_k\ k(k-1)\ldots(k-s+1) = m_{(s)}$$

de unde rezultă :

$$w(t) = \sum_{s=0}^{m} \frac{m_{(s)}}{s\,!}(t-1)^s$$

Să se demonstreze relația :

$$m_{(s)} = m_s - \frac{s(s-1)}{2\,!} m_{s-1} + \frac{s(s-1)(s-2)(3s-1)}{4\,!} m_{s-2} + \ldots$$
$$+ (-1)^{(s-1)}_{(s-1)} m_1$$

unde m_s sînt momentele inițiale ale variabilei x.

Capitolul VII

DISTRIBUȚIA NORMALĂ MULTIDIMENSIONALĂ ȘI DISTRIBUȚIILE FUNCȚIILOR DE VARIABILE ALEATORII

§ 1. Noțiuni sumare asupra formelor lineare și pătratice

Polinomul omogen și de gradul întîi $\varphi(x_1, x_2) = ax_1 + bx_2$ se numește *formă liniară* în variabilele x_1 și x_2. Numerele a și b se numesc coeficienții formei.

Polinomul omogen de gradul al doilea $\psi(x_1, x_2) = ax_1^2 + 2bx_1, x_2 + cx_2^2$ *se numește formă pătratică în variabilele* x_1 și x_2.

Acestei forme pătratice i se asociază formele lineare

$$\frac{1}{2}\frac{\partial}{\partial x_1}\Psi(x_1, x_2) = ax_1 + bx_2$$

$$\frac{1}{2}\frac{\partial}{\partial x_2}(x_1, x_2) = bx_1 + cx_2$$

Determinantul acestor două forme :

$$\Delta = \begin{vmatrix} a & b \\ b & c \end{vmatrix} = ac - b^2$$

se numește *discriminantul formei* $\Psi(x_1, x_2)$. El este un determinant *simetric*, adică are elementele simetrice față de diagonala principală, egale între ele. Forma. $\Psi(x_1, x_2)$ se poate scrie dedublînd termenii :
$\Psi(x_1, x_2) = ax_1 x_1 + bx_1 x_2 + bx_2 x_1 + cx_2 x_2$. Dacă produsului

$x_k\, x_h$ $(k, h = 1, 2)$ îi dăm drept coeficient numărul a_{kh}, atunci forma pătratică se poate scrie astfel :

$$\Psi(x_1, x_2) = a_{11} x_1^2 + a_{12} x_1 x_2 + a_{24} x_2 x_1 + a_{22} x_2^2 \qquad (1)$$

unde trebuie să luăm :

$$a_{12} = a_{21} \qquad (1_1)$$

În acest caz, formele lineare asociate formei pătratice (1) sînt :

$$\frac{1}{2}\frac{\partial \psi}{\partial x_1} = a_{11} x_1 + a_{12} x_2$$

$$\frac{1}{2}\frac{\partial \psi}{\partial x_2} = a_{21} x_1 + a_{22} x_2 \qquad (1_2)$$

iar discriminantul formei $\psi(x_1, x_2)$ va fi :

$$|\, a_{kh} \,| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \qquad (1_3)$$

Forma $\psi(x_1, x_2)$ se descompune într-o sumă de cel mult două pătrate. În acest scop, folosind relația (1_1), vom scrie forma (1) astfel :

$$\psi(x_1, x_2) = a_{11} x_1^2 + 2\, a_{12} x_1 x_2 + a_{22} x_2^2 \qquad (1_4)$$

sau ceea ce este același lucru

$$\psi(x_1 x_2) = a_{11} x_1^2 + 2\, a_{21} x_1 x_2 + a_{22} x_2^2 \qquad (1'_4)$$

Discriminantul formei are valoarea :

$$|a_{kh}| = a_{11} a_{22} - a_{12}^2 = a_{11} a_{22} - a_{21}^2 \qquad (1'_3)$$

Vom scrie acum succesiv :

$$a_{11} x_1^2 + 2a_{12} x_1 x_2 + a_{22} x_2^2 = \frac{1}{a_{11}} \Big[a_{11}^2 x_1^2 + 2\, a_{11} a_{12} x_1 x_2 +$$

$$a_{12}^2 x_2^2 + a_{11} a_{22} x_2^2 - a_{12}^2 x_2^2 \Big] = \frac{1}{a_{11}} \Big(a_{11} x_1 + a_{12} x_2 \Big)^2 + \frac{1}{a_{11}}$$

$$\Big(a_{11} a_{22} - a_{12}^2 \Big) x_2^2$$

Avem deci descompunerea :

$$a_{11} x_1^2 + 2\, a_{12} x_1 x_2 + a_{22} x_2^2 = \left(\frac{a_{11} x_1 + a_{12} x_2}{\sqrt{a_{11}}} \right)^2 + \left(\frac{\sqrt{|\, a_{kh} \,|}}{\sqrt{a_{11}}}\, x_2 \right)^2$$

Scanned with OKEN Scanner

Formele (1) sau (1_4) sau $(1'_4)$ se pot scrie prescurtat astfel :

$$\psi(x_1, x_2) = \sum_{k=1}^{2} \sum_{h=1}^{2} a_{kh}\, x_k\, x_h$$

Forma pătratică cu s variabile se definește prin relația :

$$\psi(x_1, x_2, \ldots, x_s) = a_{11}\, x_1^2 + a_{22}\, x_2^2 + \ldots + a_{ss}\, x_s^2 + 2\, a_{12}\, x_1\, x_2 + \ldots + 2\, a_{s-1,}\, x_{s-1}\, x_s \qquad (1_6)$$

sau prescurtat :

$$\psi(x_1, x_2, \ldots x_s) = \sum_{k=1}^{s} \sum_{h=1}^{s} a_{kh}\, x_k\, x_h \qquad (1'_6)$$

Discriminantul formei este determinantul simetric de ordinul s :

$$|\, a_{kh}\,| = \begin{vmatrix} a_{11} & a_{12} \ldots a_1 \\ a_{21} & a_{22} \ldots a_2 \\ \cdot & \cdot \quad \cdot \\ \cdot & \cdot \quad \cdot \\ \cdot & \cdot \quad \cdot \\ a_{s1} & a_{s2} \quad a_{ss} \end{vmatrix} \qquad (1_7)$$

unde :

$$a_{pq} = q_{qp},\ p,\ q = 1,\ 2,\ \ldots\ s \qquad (1_8)$$

Forma $(1'_6)$ se poate descompune într-o sumă de s pătrate.

Definiție. O formă pătratică se numește *pozitiv definită* dacă pentru orice sistem de valori $(x_1, x_2, \ldots x_n)\, \psi(x_1 \ldots x_n) \geqslant 0$. Pentru ca forma (1_6) să fie pozitiv definită este necesar și suficient să avem îndeplinite condițiile :

$$a_{11} > 0,\ \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0,\ \ldots \begin{vmatrix} a_{11} & a_{12} \ldots a_{1\,n} \\ a_{21} & a_{22} \ldots a_{2\,n} \\ \cdot & \cdot \quad \cdot \\ \cdot & \cdot \quad \cdot \\ a_{n1} & a_{n2} \ldots a_{nn} \end{vmatrix} > 0 \qquad (1_8)$$

În paragrafele care urmează vom considera numai forme pătratice pozitiv definite.

M a t r i c i. Tabloul format din mn elemente aşezate pe m linii şi n coloane

$$\| a_{kh} \| = \begin{Vmatrix} a_{11} & a_{12} \ldots a_{1n} \\ a_{21} & a_{22} \ldots a_{2n} \\ \cdot & \cdot \quad \cdot \\ \cdot & \cdot \quad \cdot \\ \cdot & \cdot \quad \cdot \\ a_{m1} & a_{m2} \quad a_{mn} \end{Vmatrix} \tag{1_9}$$

pentru care se definesc anumite reguli de calcul, se numeşte *matrice*. Primul indice arată linia în care se găseşte un element al matricei, iar al doilea coloana; elementul a_{34} de pildă se găseşte la intersecţia liniei a treia cu coloana a patra.

În particular, cînd m = n obţinem matricea pătrată:

$$\| a_{kh} \| = \begin{Vmatrix} a_{11} & a_{12} \ldots a_{1n} \\ a_{21} & a_{22} \ldots a_{2n} \\ \cdot & \cdot \quad \cdot \\ \cdot & \cdot \quad \cdot \\ \cdot & \cdot \quad \cdot \\ a_{n1} & a_{n2} \quad a_{nn} \end{Vmatrix} \tag{$1_9'$}$$

Matricea (1_9) se scrie de obicei prescurtat astfel: $\| a_{kh} \|$, $k = 1, 2 \ldots m$; $h = 1, 2 \ldots n$, iar matricea pătrată $(1_9')$ se notează: $\| a_{kh} \|$, $k, h = 1, 2 \ldots n$.

Determinantul de ordinul n.

$$| a_{kh} | = \begin{vmatrix} a_{11} & a_{12} \ldots a_{1n} \\ a_{21} & a_{22} \ldots a_{2n} \\ \cdot & \cdot \quad \cdot \\ \cdot & \cdot \quad \cdot \\ \cdot & \cdot \quad \cdot \\ a_{n1} & a_{n2} \quad a_{nn} \end{vmatrix} \tag{1_{10}}$$

sau scris prescurtat $| a_{kh} |$, $k, h = 1, 2 \ldots n$ i se spune de obicei determinantul matricei pătrate $\| a_{kh} \|$.

Nu vom face aici un studiu al determinanţilor sau matricelor; vom indica numai unele rezultate de care avem nevoie în dezvoltările care urmează.

Calculul determinantului de ordinul n se reduce la calculul unui determinant de ordinul (n — 1). Să considerăm determinan-

tul (1_{10}) în care să înmulțim linia întîi succesiv prin $\frac{a_{k1}}{a_{11}}$ și o scădem din linia k, (k = 2, 3 . . n). Se obține astfel :

$$|a_{kh}| = \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} - \frac{a_{21}a_{11}}{a_{11}} & \cdots & a_{2n} - \frac{a_{21}a_{1n}}{a_{11}} \\ \vdots & & \vdots \\ a_{n1} - \frac{a_{n1}a_{11}}{a_{11}} & \cdots & a_{nn} - \frac{a_{n1}a_{1n}}{a_{11}} \end{vmatrix} =$$

$$= \begin{vmatrix} a_{11} & a_{12} \cdots a_{1n} \\ 0 & a^{(1)}_{22} \cdots a^{(1)}_{2n} \\ \vdots & \vdots \\ 0 & a^{(1)}_{n2} \cdots a^{(1)}_{nn} \end{vmatrix} = a_{11} \left| a^{(1)}_{kh} \right| k, h, = 2 \ldots n$$

unde :

$$\left| a^{(1)}_{kh} \right| = \begin{vmatrix} a^{(1)}_{22} & \cdots & a^{(1)}_{2n} \\ \vdots & & \\ a^{(1)}_{n2} & & a^{(1)}_{nn} \end{vmatrix} \qquad (1_{11})$$

iar

$$a^{(1)}_{kh} = a_{kh} - \frac{a_{k1}\,a_{1h}}{a_{11}}, (k,\ h = 2,\ 3, \ldots n) \qquad (1_{12})$$

Determinantul (1_{11}) este de ordinul (n—1) și elementele sale se calculează cu formula (1_{12}). Așa de exemplu $a^{(1)}_{43}$ este :

$$a^{(1)}_{43} = a_{43} - \frac{a_{41}\,a_{13}}{a_{11}}$$

Determinantului $|a^{(1)}_{kh}|$ îi aplicăm același procedeu și obținem :

$$\left| a^{(1)}_{kh} \right| = a^{(1)}_{22} \left| a^{(2)}_{kh} \right|, \ k, h = 3,\ 4,\ \ldots\ n$$

Scanned with OKEN Scanner

unde :

$$a_{kh}^{(2)} = a_{kh}^{(1)} - \frac{a_{k1}^{(1)}\, a_{1h}^{(1)}}{a_{22}^{(1)}} \qquad (1'_{12})$$

Determinantul $|\, a_{kh} \,|$ de ordinul n se poate scrie acum :

$$|\, a_{kh} \,| = a_{11} \cdot a_{22}^{(1)} \Big|\, a_{kh}^{(2)} \,\Big|_{k,\,h=3,\,4,\,\ldots n}^{k,\,h=1,\,2,\,3\ldots n}$$

Repetînd procedeul de mai sus de (n—1) ori, din aproape în aproape obținem :

$$|\, a_{kh} \,| = a_{11}\, a_{22}^{(1)}\, a_{33}^{(2)} \ldots a_{nn}^{(n-1)} \qquad (1_{13})$$

unde.

$$a_{nn}^{(n-1)} = a_{nn}^{(n-2)} - \frac{a_{n,\,n-1}^{(n-2)}\, a_{n-1,\,n}^{(n-2)}}{a_{n-1,\,n-1}^{(n-2)}} \qquad (1''_{12})$$

Se știe că determinantul de ordinul (n—1) obținut din $|\, a_{kh} \,|$, k, h = 1,2..n, prin suprimarea liniei k și a coloanei h se numește *minorul elementului* a_{kh} ; îl vom nota prin $|\, A_{kh} \,|$. Minorul elementului a_{11}, de pildă, este :

$$|\, A_{11} \,| = \begin{vmatrix} a_{22} & \ldots & a_{2\,n} \\ a_{32} & \ldots & a_{3\,n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ a_{n\,2} & & a_{n\,n} \end{vmatrix}$$

$$\text{Matricea } E = \begin{Vmatrix} 1 & 0 & 0 \ldots 0 \\ 0 & 1 & 0 \ldots 0 \\ 0 & 0 & 1 \ldots 0 \\ \cdot & \cdot & \cdot \ldots \cdot \\ \cdot & \cdot & \cdot \ldots \cdot \\ 0 & 0 & 0 \ldots 1 \end{Vmatrix}$$

ale cărei elemente situate pe diagonala principală sînt egale cu 1 și celelalte cu zero, se numește *matricea unitate*. Se notează prescurtat astfel :

$$E = \| \delta_{kh} \| \qquad (1'_{14})$$

Scanned with OKEN Scanner

unde :

$$\delta_{kh} = \begin{cases} 1, & \text{pentru } k = h \\ 0, & \text{,, } k \neq h \end{cases} \qquad (1''_{14})$$

Prin definiție, *produsul liniei pe coloane* al matricilor pătrate

$$A = \| a_{kh} \| = \begin{Vmatrix} a_{11} & a_{12} \dots a_{1n} \\ a_{21} & a_{22} \dots a_{2n} \\ \dots & \dots \\ \dots & \dots \\ a_{n1} & a_{n2} \dots a_{nn} \end{Vmatrix}$$

$$X = \| x_{kh} \| = \begin{Vmatrix} x_{11} & x_{12} \dots x_{1n} \\ x_{21} & x_{22} \dots x_{2n} \\ \dots & \dots \\ \dots & \dots \\ x_{n1} & x_{n2} \dots x_{nn} \end{Vmatrix}$$

este matricea

$$A\,X = \| b_{kh} \| = \begin{Vmatrix} b_{11} & b_{12} \dots b_{1n} \\ b_{21} & b_{22} \dots b_{2n} \\ \dots & \dots \\ \dots & \dots \\ b_{n1} & b_{n2} \dots b_{nn} \end{Vmatrix} \qquad (1_{15})$$

în care elementul b_{kh} din linia k și coloana h este egal cu suma produselor dintre elementele liniei k din $\| a_{kh} \|$ și elementele corespunzătoare ale coloanei h din $\| x_{kh} \|$, adică :

$$b_{kh} = a_{k1}\,x_{1h} + a_{k2}\,x_{2h} + - + a_{kn}\,x_{nh} = \sum_{r=1}^{n} a_{kr}\,x_{rh} \qquad (1''_{15})$$

pentru orice k, h = 1, 2, ...n.

Matricea $\| a_{kh} \|$ se numește *simetrică*, dacă elementele așezate simetric față de diagonala principală sînt egale între ele.

În orice matrice simetrică avem deci relațiile :

$$a_{ij} = a_{ji}$$

pentru orice i, j = 1, 2..n; așa de exemplu $a_{12} = a_{21}$, $a_{57} = a_{75} \ldots a_{22} = a_{22}$ etc.
Matricea $X = \| x_{hh} \|$ se numește *inversa* matricei $A = \| a_{kh} \|$ dacă $AX = E$
Inversa matricei $\| a_{kh} \|$ se notează și $\| a_{kL} \|^{-1}$.
Un element oarecare al produsului AX fiind : (1_{17})

$$b_{kh} = \sum_{r=1}^{n} , a_{kr} x_{rh} \text{ (formula } 1'_{15})$$

avem :

$$\sum_{r=1}^{n} a_{kr} x_{rh} = \delta_{kh}, k, h = 1, 2, \ldots n \qquad (1'_{17})$$

numerele δ_{kh} fiind date de $(1''_{14})$.

Dacă h este fix și k = 1, 2,... n avem sistemul de n ecuații lineare :

$$a_{11} x_{1h} + a_{12} x_{2h} + \ldots + a_{1n} x_{nh} = \delta_{1h}$$

$$a_{21} x_{1h} + a_{22} x_{2h} + \ldots + a_{2n} x_{nh} = \delta_{2h}$$

$$\ldots\ldots\ldots\ldots$$

$$\ldots\ldots\ldots\ldots$$

$$a_{n1} x_{1h} + a_{n2} x_{2h} + \ldots + a_{nn} x_{nh} = \delta_{nh}$$

cu necunoscutele $x_{1h}, x_{2h}, .. x_{nh}$. Matricea acestui sistem este $\| a_{hk} \|$; dacă presupunem că determinantul acestei matrici

$$| a_{kh} | \neq 0$$

$$k, h = 1, 2, .. n$$

obținem necunoscutele cu formulele lui Cramer :

$$x_{ih} = \frac{| a_{kh} |^{(i)}}{| a_{kh} |}, i = 1, 2, \ldots n.$$

unde $| a_{kh} |^{(i)}$ este determinantul $| a_{kh} |$ în care elementele coloanei i s-au înlocuit prin termenii liberi $\delta_{1h}, \delta_{2h}, .. \delta_{nh}$. Dacă facem h = 1, 2..n obținem necunoscutele :

$$x_{11}, x_{12} .. x_{1n}$$

$$x_{21}, x_{22} .. x_{2n}$$

$$\ldots\ldots$$

$$\ldots\ldots$$

$$x_{n1}, x_{n2} .. x_{nn}$$

care sînt tocmai elementele matricei X, inversa matricei A.

Scanned with OKEN Scanner

Exemplu : Să calculăm inversa matricei :

$$A = \begin{Vmatrix} 1 & 2 & -2 \\ 3 & 1 & 4 \\ 0 & 2 & -1 \end{Vmatrix}$$

Vom avea :

$$\begin{Vmatrix} 1 & 2 & -2 \\ 3 & 1 & 4 \\ 0 & 2 & -1 \end{Vmatrix} \cdot \begin{Vmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{Vmatrix} \equiv \begin{Vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{Vmatrix}$$

De unde deducem sistemele cu aceeași matrice A :

$$\begin{Vmatrix} x_{11} + 2x_{21} - 2x_{31} = 1 \\ 3x_{11} + x_{21} + 4x_{31} = 0 \\ 2x_{21} - x_{31} = 0 \end{Vmatrix} \quad \begin{Vmatrix} x_{12} + 2x_{22} - 2x_{32} = 0 \\ 3x_{12} + x_{22} + 4x_{32} = 1 \\ 2x_{22} - x_{32} = 0 \end{Vmatrix} \quad \begin{Vmatrix} x_{13} + 2x_{23} - 2x_{33} = 0 \\ 3x_{13} + x_{23} + 4x_{33} = 0 \\ 2x_{23} - x_{33} = 1 \end{Vmatrix}$$

Din primul sistem deducem : $x_{11} = \frac{3}{5}$, $x_{21} = -\frac{1}{5}$, $x_{31} = -\frac{2}{5}$

din al doilea : $x_{12} = \frac{2}{15}$, $x_{22} = \frac{1}{15}$, $x_{32} = \frac{2}{15}$ și din al treilea : $x_{13} = -2$, $x_{23} = 2$, $x_{33} = 1$

Prin urmare inversa matricei A este :

$$X = \begin{Vmatrix} \frac{3}{5} & \frac{2}{15} & -2 \\ -\frac{1}{5} & \frac{1}{15} & 2 \\ -\frac{2}{5} & \frac{2}{15} & 1 \end{Vmatrix}$$

Să presupunem că matricea $A = \| a_{kh} \|$ este simetrică.
Atunci putem scrie :

$$\| a_{kh} \| = a_{11} \| a_{kh}^{(1)} \|$$

$$k, h = 1, 2, \ldots n \qquad k, h = 2, 3, \ldots n \qquad (1'_{18})$$

așa după cum am arătat mai înainte. Matricea fiind simetrică în particular, $a_{k1} = a_{1k}$ deci :

$$a_{kh}^{(1)} = a_{kh} - \frac{a_{k1}\, a_{1h}}{a_{11}} = a_{kh} - \frac{a_{1k}\, a_{1h}}{a_{11}} \qquad (1_{18})$$

Determinantul de ordinul (n — 1)

$$\left| a_{kh}^{(1)} \right| = \frac{\left| a_{kh} \right|}{a_{11}} \qquad (1_{19})$$

$$k, h = 2, 3, \ldots n$$

se obține după următoarea regulă :

1°. Se calculează matricea inversă $\| a_{kh} \|^{-1}$ în care se suprimă prima linie și prima coloană ; se obține o nouă matrice pătrată $\| a'_{kh} \|$, k, h = 2,3, .. n.

2°. Se calculează apoi inversa acestei matrici adică $\| a'_{kh} \|^{-1}$ și se obține determinantul (de ordinul n — 1) :

$$\left| a'_{kh} \right| = \left| a_{kh}^{(1)} \right|, \; k, h = 2, 3, \ldots n \qquad (1'_{19})$$

Demonstrație. Deoarece matricea $\| a_{kh} \|$ este simetrică și inversa sa $\left\| a_{kh} \right\|^{-1}$ este simetrică, deci în relația $\sum_{r=1}^{n} a_{kr} x_{rh} = \delta kh$ putem schimba k cu r în a_{kr} sau r cu h în x_{rh}, fără ca relația să se schimbe.

Va fi de ajuns să demonstrăm că inversa matricei : $a_{kh}^{(1)} \|$ este tocmai matricea obținută din $\| a_{kh} \|^{-1}$ suprimînd prima linie și prima coloană, adică să stabilim relația :

$$\sum_{r=2}^{n} a_{kr}^{(1)} x_{rh} = \delta_{kh}, \; k, h = 2, 3, \ldots n \qquad (1''_{19})$$

deoarece partea întîi reprezintă elementul din linia k și coloana h a produsului $\| a_{kh}^{(1)} \| \; \| x_{kh} \|$, k, h = 2, 3, .. n.

Vom putea scrie succesiv :

$$\sum_{r=2}^{n} a_{kr}^{(1)} x_{rh} = \sum_{r=2}^{n} \left(a_{kr} - \frac{a_{1k} a_{1r}}{a_{11}} \right) x_{rh} = \sum_{r=2}^{n} a_{kr} x_{rh} - \frac{a_{1k}}{a_{11}} \sum_{r=2}^{n} a_{1r} x_{rh} \qquad (1'''_{19})$$

Avem însă relația :

$$\sum_{r=1}^{n} a_{kr} x_{rh} = a_{k1} x_{1h} + \sum_{r=2}^{n} a_{kr} x_{rh} = \delta_{kh}$$

de unde :

$$\sum_{r=2}^{n} a_{kr} x_{rh} = \delta_{kh} - a_{k1} x_{1h} = \delta_{kh} - a_{1k} x_{1h}$$

Scanned with OKEN Scanner

și în mod analog,

$$\sum_{r=2}^{n} a_{1r} x_{rh} = \delta_{1h} - a_{11} x_{1h} = a_{11} x_{1h}$$

deoarece $h = 2, 3, \ldots n$, deci $\delta_{1h} = 0$ căci $h \neq 1$

Înlocuind aceste valori în $(1'''_{19})$ găsim :

$$\sum_{r=2}^{n} a_{kr}^{(1)} x_{rh} = \delta_{kh} - a_{k1} x_{1h} - \frac{a_{1k}}{a_{11}} (-a_{11} x_{1h}) = \delta_{kh}$$

adică relația $(1''_{19})$.

În concluzie, pentru a calcula raportul $\frac{|a_{kh}|}{a_{11}} = |a_{kh}^{(1)}|$ vom calcula matricea $|a_{kh}|^{-1}$, inversa matricei $|a_{kh}|$, în care suprimăm prima linie și prima coloană ; determinantul inversei matricei astfel obținută este raportul căutat. Această teoremă ne va folosi mai departe.

§ 2. Distribuția normală bidimensională

Distribuția normală bidimensională este caracterizată prin densitatea :

$$f(x_1, x_2) = \frac{\sqrt{|a_{kh}|}}{2\pi} e^{-\frac{I}{2} \sum_{k=1}^{2} \sum_{h=1}^{2} a_{kh} (x_k - m_k)(x_h - m_h)} \qquad (2)$$

unde suma dublă de la exponent este o formă pătratică în variabilele $(x_k - m_k)$ și $(x_h - m_h)$ pozitiv definită. Scriind dezvoltat avem :

$$\sum_{k=1}^{2} \sum_{h=1}^{2} a_{kh} (x_k - m_k)(x_h - m_h) = a_{11} (x_1 - m_1)^2 + 2 a_{12} (x_1 - m_1) (x_2 - m_2) + a_{22} (x_2 - m_2)^2 \qquad (2_1)$$

unde : $a_{11} > 0 \; |a_{kh}| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11} a_{22} - a_{12}^2 > 0 \qquad (2_2)$

Densitatea (1) formează o familie cu 5 parametrii m_1, m_2, a_{11}, $a_{12} = a_{21}$, a_{22}

Pentru simplificare vom face în (2) schimbarea de variabile :

$$\xi_1 = x_1 - m_1$$
$$\xi_2 = x_2 - m_2 \qquad (2_3)$$

Atunci funcția (2) devine :

$$\Phi(\xi_1, \xi_2) = \frac{\sqrt{|a_{kh}|}}{2\pi} e^{-\frac{1}{2}\sum_{k=1}^{2}\sum_{h=1}^{2} a_{kh}\xi_k\xi_h} \qquad (2')$$

Va trebui să demonstrăm că $\Phi(\xi_1, \xi_2)$ reprezintă o densitate de probabilitate, adică sînt îndeplinite condițiile :

$$\Phi(\xi_1, \xi_2) \geqslant 0, \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \Phi(\xi_1, \xi_2)\, d\xi_1\, d\xi_2 = 1$$

În baza relațiilor (2_2) prima condiție este evidentă. În ce privește condiția a doua, descompunînd forma pătratică din exponent după relația (1_5) obținem :

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{\sqrt{|a_{kh}|}}{2\pi} e^{-\frac{1}{2}\sum_{k=1}^{2}\sum_{h=1}^{2} a_{kh}\xi_1\xi_2}\, d\xi_1\, d\xi_2 =$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{\sqrt{|a_{kh}|}}{2\pi} e^{-\frac{1}{2}\left(\frac{a_{11}\xi_1 + a_{12}\xi_2}{\sqrt{a_{11}}}\right)^2 - \frac{1}{2}\left(\frac{\sqrt{|a_{kh}|}}{\sqrt{a_{11}}}\xi_2\right)^2} d\xi_1, d\xi_2$$

Facem schimbarea de variabile :

$$\frac{a_{11}}{\sqrt{a_{11}}}\xi_1 + \frac{a_{12}}{\sqrt{a_{11}}}\xi_2 = z_1, \sqrt{\frac{|a_{kh}|}{a_{11}}}\xi_2 = z_2 \qquad (2_4)$$

De unde deducem :

$$\xi_1 = \frac{1}{\sqrt{a_{11}}} z_1 - \frac{a_{12}}{\sqrt{a_{11}|a_{kh}|}} z_2, \xi_2 = \sqrt{\frac{a_{11}}{|a_{kh}|}} z_2 \qquad (2'_4)$$

Determinantul funcțional al transformării este :

$$\frac{D(\xi_1, \xi_2)}{D(z_1, z_2)} = \begin{vmatrix} \frac{\partial \xi_1}{\partial z_1} & \frac{\partial \xi_1}{\partial z_2} \\ \frac{\partial \xi_2}{\partial z_1} & \frac{\partial \xi_2}{\partial z_2} \end{vmatrix} = \begin{vmatrix} \frac{1}{\sqrt{a_{11}}} & -\frac{a_{12}}{\sqrt{a_{11}|a_{kh}|}} \\ 0 & \sqrt{\frac{a_{11}}{|a_{kh}|}} \end{vmatrix} = \frac{1}{\sqrt{|a_{kh}|}}$$

Scanned with OKEN Scanner

Vom putea deci scrie integrala dublă :

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\Phi(\xi_1,\xi_2)\,d\xi_1\,d\xi_2=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{\sqrt{|a_{kh}|}}{2\pi}e^{-\frac{1}{2}z_1^2}e^{-\frac{1}{2}z_2^2}$$

$$\frac{1}{\sqrt{|a_{kh}|}}dz_1\,dz_2=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z_1^2}dz_1\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z_2^2}dz_2=1$$

deoarece știm că integrala :

$$\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}dx=1.$$

Așadar, $\Phi(\xi_1, \xi_2)$ reprezintă o densitate de probabilitate. Probabilitatea ca punctul $M(\xi_1, \xi_2)$ să se găsească într-un domeniu oarecare din planul $\xi_1\ \xi_2$ se obține integrînd funcția de densitate $\Phi(\xi_1, \xi_2)$ în acel domeniu. Geometric, funcția $\Phi(\xi_1, \xi_2)$ reprezintă față de un sistem de axe tridreptunghic $0\ \xi_1\ \xi_2\ \xi_3$ o suprafață în formă de cupolă situată deasupra planului $\xi_1\ \xi_2$.

Extremele acestei funcții se obțin din sistemul :

$$\frac{\partial}{\partial\xi_1}\Phi(\xi_1,\xi_2)=0,\quad\frac{\partial}{\partial\xi_2}\Phi(\xi_1,\xi_2)=0$$

Scriind $\Phi(\xi_1, \xi_2)$ sub formă dezvoltată avem :

$$\Phi(\xi_1,\xi_2)=\frac{\sqrt{|a_{kh}|}}{2\pi}e^{-\frac{1}{2}(a_{11}\xi_1^2+2a_{12}\xi_1\xi_2+a_{22}\xi_2^2)}$$

vom găsi derivatele parțiale :

$$\frac{\partial}{\partial\xi_1}\Phi(\xi_1,\xi_2)=-(a_{11}\xi_1+a_{12}\xi_2)\,\Phi(\xi_1,\xi_2)$$

$$\frac{\partial}{\partial\xi_2}\Phi(\xi_1,\xi_2)=-(a_{21}\xi_1+a_{22}\xi_2)\,\Phi(\xi_1,\xi_2)\qquad(2_5)$$

Avem deci sistemul omogen :

$$a_{11}\xi_1+a_{12}\xi_2=0$$
$$a_{21}\xi_1+a_{22}\xi_2=0\qquad(2_6)$$

Deoarece determinantul acestui sistem este :
$|a_{kh}| = a_{11}\,a_{22} - a_{12}^2 \neq 0$ sistemul admite numai soluția nulă $\xi_1 = \xi_2 = 0$ Așadar,

$$\frac{\partial}{\partial \xi_1}\Phi(0,\ 0) = 0, \quad \frac{\partial}{\partial \xi_2}\Phi(0,\ 0) = 0 \qquad (2_7)$$

Vom cerceta acum semnul expresiei :

$$\Delta(\xi_1, \xi_2) = \frac{\partial^2 \Phi}{\partial \xi_1^2} \cdot \frac{\partial^2 \Phi}{\partial \xi_2^2} - \left(\frac{\partial^2 \Phi}{\partial \xi_1\, \partial \xi_2}\right)^2$$

pentru $\xi_1 = \xi_2 = 0$
Din relația (2_5) deducem :

$$\frac{\partial^2}{\partial \xi_1^2}\Phi(\xi_1, \xi_2) = (a_{11}\xi_1 + a_{12}\xi_2)^2\,\Phi(\xi_1, \xi_2) - a_{11}\,\Phi(\xi_1, \xi_2)$$

$$\frac{\partial^2}{\partial \xi_1,\, \partial \xi_2}\Phi(\xi_1, \xi_2) = (a_{11}\xi_1 + a_{12}\xi_2)(a_{21}\xi_1 + a_{22}\xi_2)\,\Phi\xi_1, \xi_2) - a_{21}\,\Phi(\xi_1, \xi_2)$$

$$\frac{\partial^2}{\partial \xi_2^2}\Phi(\xi_1, \xi_2) = (a_{21}\xi_1 + a_{22}\xi_2)\,\Phi(\xi_1, \xi_2) - a_{22}\,\Phi(\xi_1, \xi_2)$$

De unde rezultă :

$$\frac{\partial^2}{\partial \xi_1^2}\Phi(0,0) = -a_{11}; \quad \frac{\partial^2}{\partial \xi_1\, \partial \xi_2}\Phi(0,0) = -a_{12}; \quad \frac{\partial^2}{\partial \xi_2^2}\Phi(0,0) = -a_{22}$$

Înlocuind în $\Delta(\xi_1, \xi_2)$ obținem : $\Delta(\xi_1, \xi_2) = a_{11}\,a_{22} - a_{12}^2 = |a_{k,h}| > 0$
Funcția $\Phi(\xi_1, \xi_2)$ are deci un extrem pentru $\xi_1 = \xi_2 = 0$
Deoarece :

$$\frac{\partial^2}{\partial \xi_1^2}\Phi(0,0) = -a_{11} < 0$$

rezultă că în punctul de coordonate $\xi_1 = \xi_2 = 0$ suprafața are un maxim egal cu $\Phi(0, 0) = \sqrt{|a_{kh}|} : 2\pi$.
Avem însă din (2_1)

$$\xi_1 = x_1 - m_1 = 0, \qquad \xi_2 = x_2 - m_2 = 0$$

deci funcția $f(x_1, x_2)$ are în punctul de coordonate (m_1, m_2) un maxim egal cu :

$$f(m_1, m_2) = \frac{\sqrt{|a_{kh}|}}{2\pi}$$

Scanned with OKEN Scanner

Dacă intersectăm suprafața cu planul paralel cu $x_1\ x_2$ de ecuație

$$x_3 = c < \frac{\sqrt{|a_{kh}|}}{2\pi}$$

obținem elipsa $a_{11}\ (x_1 - m_1)^2 + 2\ a_{12}\ (x_1 - m_1)\ (x_2 - m_2) + + a_{22}\ (x_2 - m_2)^2 = 1n\ \frac{|a_{kh}|}{(2\ \pi c)^2}$ cu centrul în punctul (m_1, m_2).

Orice plan perpendicular pe planul $x_1\ x_2$ taie suprafața după o curbă normală. În graficul alăturat jos am reprezentat porțiunea din suprafața $f\ (x_1,\ x_2)$, situată deasupra planului $x_3 = c$

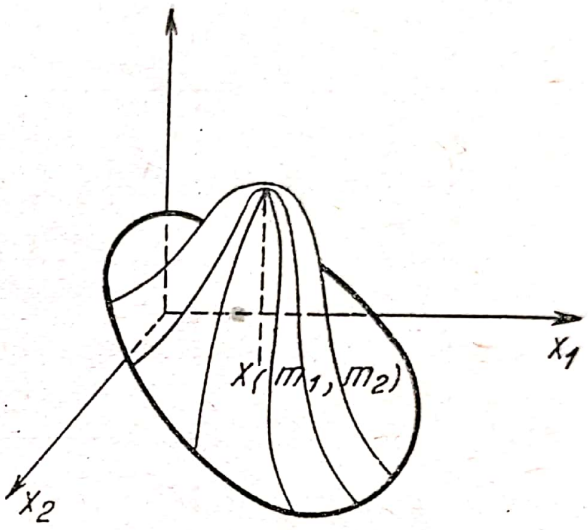

Fig. 42

Momentele distribuției. Vom calcula în acest scop funcția generatoare de momente:

$$m\ (t_1,\ t_2) = M\ (e^{\ t_1\ \xi_1 + t_2\ \xi_2}) =$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\sqrt{|a_{kh}|}}{2\ \pi}\ e^{\ t_1\ \xi_1 + t_2\ \xi_2}\ e^{\ -\frac{1}{2} \sum_{k=1}^{2} \sum_{h=1}^{2} a_{kh}\ \xi_k\ \xi_h}\ d\ \xi_1\ d\ \xi_2 \qquad (3)$$

Forma pătratică de la exponent se poate descompune în sumă de pătrate, în mod succesiv astfel:

$$-\frac{1}{2}\ (a_{11}\ \xi_1^2 + 2\ a_{12}\ \xi_1\ \xi_2 + a_{22}\ \xi_2^2) + t_1\ \xi_1 + t_2\ \xi_2 =$$

$$= -\frac{1}{2\ a_{11}}\ [a_{11}^2\ \xi_1^2 + 2\ a_{11}\ a_{12}\ \xi_1\ \xi_2 + a_{11}\ a_{22}\ \xi_2^2 - 2\ a_{11}\ t_1\ \xi_1 - 2\ a_{11}\ t_2\ \xi_2] =$$

$$= -\frac{1}{2\ a_{11}}\ [a_{11}^2\ \xi_1^2 + 2\ a_{11}\ (a_{12}\ \xi_2 - t_1) + (a_{12}\ \xi_2 - t_1)^2 - (a_{12}\ \xi_2 - t_1)^2 + + a_{11}\ a_{22}\ \xi_2^2 - 2\ a_{11}\ t_2\ \xi_3] =$$

$$= -\frac{1}{2\ a_{11}}\ [(a_{11}\ \xi_1 + a_{12}\ \xi_2 - t_1)^2 + (a_{11}\ a_{22} - a_{12}^2)\ \xi_2^2 + 2\ (a_{12}\ t_1 - a_{11}\ t_2)\ \xi_2 -$$

$$- t_1^2] = -\frac{1}{2}\left(\frac{a_{11}\,\xi_1 + a_{12}\,\xi_2 - t_1}{\sqrt{a_{11}}}\right)^2 - \frac{1}{2\,a_{11}\,|a_{kh}|}\,[\,|a_{kh}|^2\,\xi_2^2 +$$

$$+ 2\,(a_{12}\,t_1 - a_{11}\,t_2)\,|a_{kh}|\,\xi_2 - |a_{kh}|\,t_1^2] =$$

$$= -\frac{1}{2}\left(\frac{a_{11}\,\xi_1 + a_{12}\,\xi_2 - t_1}{\sqrt{a_{11}}}\right)^2 - \frac{1}{2}\left[\frac{|a_{kh}|\,\xi_2 + (a_{12}\,t_1 - a_{11}\,t_2)}{\sqrt{a_{11}\,|a_{kh}|}}\right]^2 +$$

$$+ \frac{1}{2}\,\frac{a_{22}\,t_1^2 - 2\,a_{12}\,t_1\,t_2 + a_{11}\,t_2^2}{a_{11}\,|a_{kh}|} \qquad (2_3)$$

Dacă introducem în relația (3) și facem schimbarea de variabile

$$\frac{a_{11}\,\xi_1 + a_{12}\,\xi_2 - t_1}{\sqrt{a_{11}}} = z_1\,;\quad \frac{|a_{kh}|\,\xi_2 + (a_{12}\,t_1 - a_{11}\,t_2)}{\sqrt{a_{11}\,|a_{kh}|}} = z_2$$

vom găsi funcția generatoare de momente :

$$m\,(t_1, t_2) = e^{\frac{a_{22}\,t_1^2 - 2\,a_{12}\,t_1\,t_2 + a_{11}\,t_2^2}{2\,|a_{kh}|}} \qquad (3')$$

deoarece integrala dublă :

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{\sqrt{|a_{kh}|}}{2\pi}\,e^{-\frac{1}{2}z_1^2}\cdot e^{-\frac{1}{2}z_2^2}\left|\frac{D\,(\xi_1\,\xi_2)}{D\,(z_1, z_2)}\right| dz_1\,dz_2 = 1$$

Derivatele parțiale în raport cu t_1, de pildă, sînt :

$$\frac{\partial}{\partial t_1}\,m\,(t_1, t_2) = \frac{a_{22}\,t_1 - a_{12}\,t_2}{|a_{kh}|}\,m\,(t_1, t_2)$$

$$\frac{\partial^2}{\partial t_1\,\partial t_2}\,m\,(t_1, t_2) = \left(\frac{a_{22}\,t_1 - a_{12}\,t_2}{|a_{kh}|}\cdot\frac{a_{11}\,t_2 - a_{12}\,t_1}{|a_{kh}|} - \frac{a_{12}}{|a_{kh}|}\right) m\,(t_1\,t_2)$$

$$\frac{\partial^2}{\partial t_1^2}\,m\,(t_1, t_2) = \left[\left(\frac{(a_{22}\,t_1 - a_{12}\,t_2}{|a_{kh}|}\right)^2 + \frac{a_{22}}{|a_{kh}|}\right] m\,(t_1, t_2)$$

De unde rezultă momentele :

$$M\,(\xi_1) = \frac{\partial}{\partial t_1}\,m\,(0,0) = 0$$

$$M\,(\xi_1, \xi_2) = \frac{\partial^2}{\partial t_1\,\partial t_2}\,m\,(0,\,0) = -\frac{a_{12}}{|a_{kh}|} \qquad (3_1)$$

$$M\,(\xi_1^2) = \frac{\partial^2}{\partial t_1^2}\,m\,(0,\,0) = \frac{a_{22}}{|a_{kh}|}$$

Scanned with OKEN Scanner

În mod analog, obținem :

$$M(\xi_2) = 0,\ M(\xi_2^2) = \frac{a_{11}}{|a_{kh}|} \tag{3_2}$$

Întrucît : $\xi_1 = x_1 - m_1$, $\xi_2 = x_2 - m_2$, deducem :

$$M(\xi_1) = M(x_1 - m_1) = M(x_1) - m_1 = 0$$
$$M(\xi_2) = M(x_2 - m_2) = M(x_2) - m_2 = 0$$

Așadar, parametrii m_1 și m_2 sînt respectiv valorile medii ale variabilelor x_1 și x_2 :

$$m_1 = M(x_1), \qquad m_2 = M(x_2) \tag{3_3}$$

Să calculăm acum parametrii a_{11}, a_{12}, a_{22}.
Din relațiile (3_1) avem :

$M(\xi_1, \xi_2) = M(x_1 - m_1)(x_2 - m_2) = -\frac{a_{12}}{|a_{kh}|} = \rho\sigma_1\ \sigma_2$ unde ρ este corelația variabilelor x_1 și x_2, iar

$$\sigma_1^2 = M(\xi_1^2) = M(x_1 - m_1)^2 = \frac{a_{22}}{|a_{kh}|}$$

$$\sigma_2^2 = M(\xi_2^2) = M(x_2 - m_2)^2 = \frac{a_{11}}{|a_{kh}|}$$

De aici formăm sistemul în a_{11}, a_{12}, a_{22} :

$$a_{11} = \sigma_2^2 |a_{kh}|$$
$$a_{22} = \sigma_1^2 |a_{kh}|$$
$$a_{12} = -\rho\sigma_1\sigma_2 |a_{kh}|$$

din care deducem combinația :
$a_{11}\ a_{22} - a_{12}^2 = \sigma_1^2\ \sigma_2^2\ (1 - \rho^2)\ |a_{kh}|^2$ și cum $|a_{kh}| = a_{11}\ a_{22} - a_{12}^2$, rezultă :

$$|a_{kh}| = \frac{1}{\sigma_1^2\sigma_2^2(1 - \rho^2)} \tag{3_4}$$

înlocuind în sistemul de mai sus se găsește :

$$a_{11} = \frac{1}{\sigma_1^2(1 - \rho^2)}$$

$$a_{12} = \frac{-\rho}{\sigma_1\sigma_2(1 - \rho^2)} \tag{3_5}$$

$$a_{22} = \frac{1}{\sigma_2^2(1 - \rho^2)}$$

Scanned with OKEN Scanner

Densitatea (1) se poate scrie acum :

$$f(x_1, x_2) = \frac{1}{2\pi\,\sigma_1\,\sigma_2\sqrt{1-\rho^2}}\, e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-m_1}{\sigma_1}\right)^2 - 2\rho\frac{x_1-m_1}{\sigma_1}\cdot\frac{x_2-m_2}{\sigma_2} + \left(\frac{x_2-m_2}{\sigma_2}\right)^2\right]} \tag{4}$$

Cu această formulă se definește, de obicei, densitatea distribuției normale bidimensionale.

Distribuții marginale. Densitatea marginală în raport cu ξ_1 de exemplu, este :

$$\Phi_1(\xi_1) = \int_{-\infty}^{+\infty} \Phi(\xi_1, \xi_2)\, d\xi_2 = \int_{-\infty}^{+\infty} \frac{\sqrt{|a_{kh}|}}{2\pi}\, e^{-\frac{1}{2}\sum_{k=1}^{2}\sum_{h=1}^{2} a_{kh}\xi_k\xi_h}\, d\xi_2 =$$

$$= \frac{e^{-\frac{1}{2}a_{11}\xi_1^2}\sqrt{|a_{kh}|}}{2\pi}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}(a_{22}\xi_2^2 + 2a_{12}\xi_1\xi_2)}\, d\xi_2$$

Exponentul de sub integrală se poate scrie succesiv :

$$-\frac{1}{2}(a_{22}\xi_2^2 + 2a_{12}\xi_1\xi_2) = -\frac{1}{2a_{22}}[a_{22}^2\xi_2^2 + 2a_{12}a_{22}\xi_1\xi_2 + a_{12}^2\xi_1^2 -$$

$$-a_{12}^2\xi_1^2] = -\frac{1}{2}\left(\frac{a_{22}\xi_2 + a_{12}\xi_1}{\sqrt{a_{22}}}\right)^2 + \frac{1}{2}\left(\frac{(a_{12}}{\sqrt{a_{22}}}\xi_1\right)^2$$

Prin substituția :

$$\frac{a_{22}\xi_2 + a_{12}\xi_1}{\sqrt{a_{22}}} = z, \qquad d\xi_2 = \frac{1}{\sqrt{a_{22}}}\,dz$$

integrala de mai sus devine :

$$\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}(a_{22}\xi_2^2 + 2a_{12}\xi_1\xi_2)}\, d\xi_2 = \frac{e^{\frac{a_{12}^2\xi_1^2}{2a_{22}}}}{\sqrt{a_{22}}}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2}\, dz =$$

$$= \frac{e^{\frac{a_{12}^2\xi_1^2}{2a_{22}}}}{\sqrt{a_{22}}}$$

Înlocuind în densitatea marginală obținem :

$$\Phi_1(\xi_1) = \frac{\sqrt{|a_{kh}|}}{\sqrt{a_{22}}\cdot\sqrt{2\pi}}\, e^{-\frac{1}{2}(a_{11}\xi_1^2 + \frac{a_{12}^2\xi_1^2}{2a_{22}})} = \frac{1}{\sqrt{2\pi}}\sqrt{\frac{|a_{kh}|}{a_{22}}}\, e^{-\frac{1}{2}\frac{|a_{kh}|}{a_{22}}\xi_1^2}$$

Din relațiile (3_4) și (3_5) avem însă : $\frac{a_{kh}}{a_{22}} = \frac{1}{\sigma_1^2}$

așa încît, densitatea marginală a lui x_1 va fi :

$$f_1(x_1) = \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{1}{2}\left(\frac{x_1 - m_1}{\sigma_1}\right)^2} \qquad (4_1)$$

adică densitatea distribuției unidimensionale x_1.
În mod analog, pentru x_2 găsim :

$$f_2(x_2) = \frac{1}{\sqrt{2\pi}\,\sigma_2} \cdot e^{-\frac{1}{2}\left(\frac{x_2 - m_2}{\sigma_2}\right)^2}$$

Distribuții condiționate. Se știe că densitatea lui x_1 condiționată de x_2 este dată de formula :

$$f(x_1/x_2) = f_{x_2}(x_1) = \frac{f(x_1, x_2)}{f_2(x_2)}$$

Folosind pentru $f(x_1, x_2)$ forma (4), iar pentru $f_2(x_2)$ forma $(4'_1)$ obținem :

$$f_{x_2}(x_1) = \frac{1}{\sqrt{2\pi}\,\sigma_1\sqrt{1-\rho^2}} \times$$

$$\times \frac{e^{-\frac{1}{2\sigma_1^2(1-\rho^2)}\left[(x_1 - m_1)^2 - 2\rho\frac{\sigma_1}{\sigma_2}(x_1 - m_1)(x_2 - m_2) + \frac{\sigma_1^2}{\sigma_2^2}(x_2 - m_2)^2\right]}}{e^{-\frac{\sigma_1^2(1-\rho^2)(x_2 - m_2)^2}{2\sigma_2^2\sigma_2^2(1-\rho^2)}}} =$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_1\sqrt{1-\rho^2}} \times$$

$$\times e^{-\frac{1}{2\sigma_1^2(1-\rho^2)}\left[(x_1 - m_1)^2 - 2\rho\frac{\sigma_1}{\sigma_2}(x_1 - m_1)(x_2 - m_2) + \rho^2\frac{\sigma_1^2}{\sigma_2^2}(x_2 - m_2)^2\right]}$$

De unde rezultă :

$$f_{x_2}(x_1) = \frac{1}{\sqrt{2\pi}\,\sigma_1\sqrt{1-\rho^2}} e^{-\frac{1}{2\sigma_1^2(1-\rho)^2}\left[x_1 - m_1 - \rho\frac{\sigma_1}{\sigma_2}(x_2 - m_2)\right]^2} \qquad (4_2)$$

Scanned with OKEN Scanner

adică o distribuție normală unidimensională cu valoarea medie :

$$m_1 + \rho \frac{\sigma_1}{\sigma_2} (x_2 - m_2) \tag{4_3}$$

și cu dispersia $\sigma_1^2 (1 - \rho^2)$. Distribuția lui x_2 condiționată de x_1 va fi :

$$f_{x_1}(x_2) = \frac{1}{\sqrt{2\pi}\ \sigma_2 \sqrt{1 - \rho^2}} e^{-\frac{1}{2}\left[x_2 - m_2 - \rho \frac{\sigma_2}{\sigma_1}(x_1 - m_1)\right]^2} \tag{4_2'}$$

cu media : $m_2 + \rho \frac{\sigma_2}{\sigma_1} (x_1 - m_1)$ și dispersia $\sigma_2^2 (1 - \rho^2)$.

Regresiune. Media variabilei x_1, în distribuția condiționată de x_2 se numește *funcția de regresiune* a variabilei x_1. Ea depinde evident de x_2. Notînd cu $R_1 (x_2)$ după definiția egresiunii, avem :

$$R_1 (x_2) = m_1 + \rho \frac{\sigma_1}{\sigma_2} (x_2 - m_2) \tag{5}$$

În mod analog se definește funcția de regresiune a variabilei x_2. Notînd-o cu $R_2 (x_1)$ vom avea :

$$R_2 (x_1) = m_2 + \rho \frac{\sigma_2}{\sigma_1} (x_1 - m_1) \tag{5'}$$

Curbele din planul $x_1\ x_2$:

$$x_1 = R_1 (x_2), \qquad x_2 = R_2 (x_1) \tag{5_1}$$

se numesc *curbele de regresiune* respectiv ale lui x_1 și x_2.

În cazul de care ne ocupăm, curbele de regresiune sînt dreptele :

$$x_1 = m_1 + \rho \frac{\sigma_1}{\sigma_2} (x_2 - m_2)$$

$$x_2 = m_2 + \rho \frac{\sigma_2}{\sigma_1} (x_1 - m_1) \tag{5_2}$$

Curba de regresiune a lui x_1 de pildă indică variațiile valorii medii a lui x_1 în raport cu diferitele valori ale lui x_2.

Scanned with OKEN Scanner

În graficul de mai jos, am reprezentat în planul $x_1\ x_2$ linia de regresiune a lui x_2 față de un sistem de trei axe perpendiculare $0x_1$, x_2, x_3 și densitatea condiționată a lui x_2 : $x_3 = f_{x_1}(x_2)$ = curba $(4'_2)$ pentru diferite valori x'_1, x''_1, x'''_1 ale lui x_1. Aceste curbe se găsesc în planele : $x_2 = x'_1$, $x_2 = x''_1$, $x_2 = x'''_1$ paralele cu planul $x_2\ x_3$.

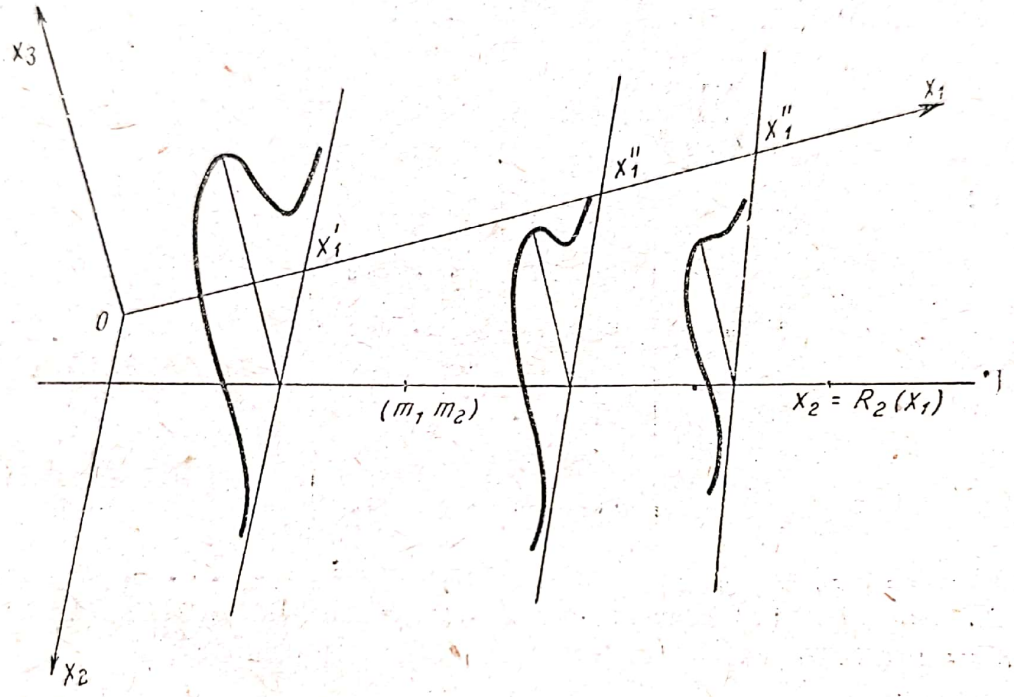

Fig. 43

Funcția de repartiție a distribuției normale bidimensionale este :

$$F(x_1, x_2) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f(s, t)\, dt\, ds =$$

$$= \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} \frac{\sqrt{|a_{kh}|}}{2\pi}\, e^{-\frac{1}{2}\sum_{k=1}^{2}\sum_{h=1}^{2} a_{kh}(x_k - m_k)(x_h - m_h)}\, dx_2\, dx_1$$

sau făcînd schimbarea (2_1) :

$$F(x_1, x_2) = F'(\xi_1, \xi_2) = \int_{-\infty}^{\xi_1}\int_{-\infty}^{\xi_2} \frac{\sqrt{|a_{kh}|}}{2\pi}\, e^{-\frac{1}{2}\sum_{k=1}^{2}\sum_{h=1}^{2} a_{kh}\xi_k\xi_h}\, d\xi_1\, d\xi_2 \qquad (6)$$

Pentru funcția $F'(\xi_1, \xi_2)$ sînt construite tabele speciale.

Distribuția normală cu s dimensiuni este caracterizată prin funcția de densitate :

$$f(x_1, x_2, \ldots, x_s) = \frac{\sqrt{|a_{kh}|}}{(2\pi)^{\frac{s}{2}}} e^{-\frac{1}{2}\sum_{k=1}^{s}\sum_{h=1}^{s} a_{kh}(x_k - m_k)(x_h - m_h)} \tag{7}$$

unde $|a_{kh}|$ este discriminantul formei pătratice pozitiv definite, în variabilele $(x_k - m_k)$, $1 \leqslant k \leqslant s$:

$$\sum_{k=1}^{s}\sum_{h=1}^{s} a_{kh}(x_k - m_k)(x_h - m_h) \tag{7_1}$$

adică :

$$|a_{kh}| = \begin{vmatrix} a_{11} & a_{12} \ldots a_{1s} \\ a_{21} & a_{22} \ldots a_{2s} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ a_{s1} & a_{s2} \ldots a_{ss} \end{vmatrix} > 0 \tag{7_2}$$

cu condiția $a_{pq} = a_{qp}$ pentru orice $p, q = 1, 2, \ldots s$.

În formula (7) făcînd $s = 2$ se obține densitatea (2). Pentru $s = 1$ avem evident densitatea unidimensională

$$f(x_1) = \frac{\sqrt{|a_{11}|}}{(2\pi)^{\frac{s}{2}}} e^{-\frac{1}{2}a_{11}(x_1 - m_1)^2} = \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{1}{2}\left(\frac{x_1 - m_1}{\sigma_1}\right)^2}$$

Funcția de s variabile $f(x_1, x_2, \ldots x_s)$ reprezintă în spațiul cu $(s + 1)$ dimensiuni o *hipersuprafață* așa după cum $f(x_1)$ reprezintă o curbă în spațiul cu două dimensiuni $x_1\ x_2$, $f(x_1, x_2)$ o suprafață în spațiul $x_1\ x_2\ x_3$ etc. Problemele tratate în cazul bidimensional se extind imediat la cazul general și se rezolvă în același mod.

Să arătăm că $f(x_1, x_2, \ldots, x_s)$ este într-adevăr o densitate de probabilitate, adică :

$$\int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} f(x_1, x_2, \ldots x_s)\, dx_1\, dx_2 \ldots dx_s = 1 \tag{7_3}$$

Mai întîi vom face schimbarea de variabile :

$$x_k - m_k = \xi_k \qquad (k = 1, 2, \ldots, s) \tag{7}$$

Scanned with OKEN Scanner

Densitatea (7) devine :

$$h(\xi_1, \xi_2, \ldots, \xi_s) = \left(\frac{1}{2\pi}\right)^{\frac{s}{2}} \sqrt{|a_{kh}|}\; e^{-\frac{1}{2}\sum_{k=1}^{s}\sum_{h=1}^{s} a_{kh}\,\xi_k\,\xi_h} \qquad (7')$$

Vom integra mai întîi în raport cu variabila ξ_1.
Forma pătratică de la exponent se poate scrie succesiv :

$$\sum_{k=1}^{s}\sum_{h=1}^{s} a_{kh}\,\xi_k\,\xi_h = a_{11}\xi_1^2 + \sum_{h=2}^{s} a_{1h}\,\xi_1\,\xi_h + \sum_{k=2}^{s} a_{k1}\,\xi_k\,\xi_1 +$$

$$+ \sum_{k=2}^{s}\sum_{h=2}^{s} a_{kh}\,\xi_k\,\xi_h = a_{11}\left(\xi_1^2 + 2\,\xi_1 \frac{1}{a_{11}}\sum_{k=2}^{s} a_{1k}\,\xi_k\right) + \sum_{k=2}^{s}\sum_{h=2}^{s} a_{kh}\,\xi_k\,\xi_h$$

matricea $\| a_{kh} \|$ fiind simetrică, adică $a_{1h} = a_{h1}$, pentru orice $h = 1, 2, \ldots, s$, deci :

$$\sum_{k=2}^{s} a_{1k}\,\xi_1\,\xi_k = \sum_{k=1}^{s} a_{k1}\,\xi_k\,\xi_1$$

Completînd pătratul din paranteză obținem :

$$\sum_{k=1}^{s}\sum_{h=1}^{s} a_{kh}\,\xi_k\,\xi_h = a_{11}\left(\xi_1 + \frac{1}{a_{11}}\sum_{k=2}^{s} a_{1k}\,\xi_k\right)^2 +$$

$$+ \sum_{k=2}^{s}\sum_{h=2}^{s} a_{kh}\,\xi_k\,\xi_h - \frac{1}{a_{11}}\left(\sum_{k=2}^{s} a_{1k}\,\xi_k\right)^2$$

Putem însă scrie relația :

$$\left(\sum_{k=2}^{s} a_{1k}\,\xi_k\right)^2 = \sum_{k=2}^{s}\sum_{h=2}^{s} a_{1k}\,a_{1h}\,\xi_k\,\xi_h$$

așa încît relația precedentă se poate scrie :

$$\sum_{k=1}^{s}\sum_{h=1}^{s} a_{kh}\,\xi_k\,\xi_h = a_{11}\left(\xi_1 + \frac{1}{a_{11}}\sum_{k=2}^{s} a_{1k}\,\xi_k\right)^2 +$$

$$+ \sum_{k=2}^{s}\sum_{h=2}^{s}\left(a_{kh} - \frac{a_{1k}\,a_{1h}}{a_{11}}\right)\xi_k\,\xi_h$$

Prin substituția :

$$u_1 = \xi_1 + \frac{1}{a_{11}}\sum_{k=2}^{s} a_{1k}\,\xi_k$$

$$a_{kh}^{(1)} = a_{kh} - \frac{a_{1k}\,a_{1h}}{a_{11}}, \qquad k, h = 2, 3, \ldots, s \qquad (7_5$$

Scanned with OKEN Scanner

se găsește :

$$\sum_{k=1}^{s}\sum_{h=1}^{s} a_{kh}\,\xi_k\,\xi_h = a_{11}\,u_1^2 + \sum_{k=2}^{s}\sum_{h=2}^{s} a_{kh}^{(1)}\,\xi_k\,\xi_h$$

Integrînd în raport cu ξ_1 obținem :

$$\int_{-\infty}^{+\infty} h(\xi_1, \xi_2 \ldots, \xi_s)\,d\xi_1 = \int_{-\infty}^{+\infty} \left(\frac{1}{2\pi}\right)^{\frac{s}{2}} \sqrt{|a_{kh}|}\; e^{-\frac{1}{2}\sum_{k=1}^{s}\sum_{h=1}^{s} a_{kh}\xi_k\xi_h}\,d\xi_1 =$$

$$= \int_{-\infty}^{+\infty} \left(\frac{1}{2\pi}\right)^{\frac{s}{2}} \sqrt{|a_{kh}|}\; e^{-\frac{1}{2}a_{11}u_1^2 - \frac{1}{2}\sum_{k=2}^{s}\sum_{h=2}^{s} a_{kh}^{(1)}\xi_k\xi_h}\,du_1$$

Facem din nou schimbarea :

$$u_1\sqrt{a_{11}} = v_1, \quad du_1 = \frac{1}{\sqrt{a_{11}}}\,dv_1$$

de unde deducem :

$$\int_{-\infty}^{+\infty} h(\xi_1, \xi_2, \ldots \xi_s)\,d\xi_1 = \left(\frac{1}{2\pi}\right)^{\frac{s-1}{2}} \frac{\sqrt{|a_{kh}|}}{\sqrt{a_{11}}}\; e^{-\frac{1}{2}\sum_{k=2}^{s}\sum_{h=2}^{s} a_{kh}^{(1)}\xi_k\xi_h}$$

deoarece :

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}v_1^2}\,dv_1 = 1$$

Dacă vom nota :

$$h_1(\xi_2, \ldots \xi_n) = \int_{-\infty}^{+\infty} h(\xi_1, \xi_2, \ldots \xi_s)\,d\xi_1$$

o nouă integrare în raport cu ξ_2 ne conduce la rezultatul :

$$h_2(\xi_3, \ldots \xi_s) = \int_{-\infty}^{+\infty} h_1(\xi_2, \xi_3, \ldots \xi_s)\,d\xi_2 = \left(\frac{1}{2\pi}\right)^{\frac{s-2}{2}}$$

$$\frac{\sqrt{|a_{kh}|}}{\sqrt{a_{11}}\cdot\sqrt{a_{22}^{(1)}}}\; e^{-\frac{1}{2}\sum_{k=3}^{s}\sum_{h=3}^{s} a_{kh}^{(2)}\xi_k\xi_h}$$

unde am notat :

$$a_{kh}^{2} = a_{kh}^{(1)} - \frac{a_{1k}^{(1)}\,a_{1h}^{(1)}}{a_{22}^{(1)}}, \quad k, h = 3, 4, \ldots, s \qquad (7'_5)$$

Scanned with OKEN Scanner

După (s—1) integrări obținem evident valoarea :

$$h_{s-1}(\xi_s) = \int_{-\infty}^{+\infty} h_{s-2}(\xi_{s-1}, \xi_s)\, d\xi_{s-1} = \left(\frac{1}{2\pi}\right)^{\frac{1}{2}} \frac{\sqrt{|a_{kh}|}}{\sqrt{a_{11} \cdot a_{22}^{(1)} \ldots a_{s-1,\,s-1}^{(s-2)}}} \cdot e^{-\frac{1}{2} a_{ss}^{(s-1)} \xi_s^2} \qquad (7_6)$$

unde :

$$a_{ss}^{(s-1)} = a_{s-1,\,s-1}^{(s-2)} - \frac{a_{1,\,s-1}^{(s-2)} \cdot a_{1,\,s-1}^{(s-2}}{a_{s-1,\,s-1}^{(s-2}} \qquad (7_5'')$$

o ultimă integrare dă prin substituția :

$$\xi_1 \sqrt{a_{ss}^{(s-1)}} = v_s,\ d\xi_s = \frac{dv_s}{\sqrt{a_{ss}^{(s-1)}}}$$

$$h_n = \int_{-\infty}^{+\infty} h_{n-1}(\xi_s)\, d\xi_s = \frac{\sqrt{|a_{kh}|}}{\sqrt{a_{11}\, a_{22}^{(1)}\, a_{33}^{(2)} \ldots a_{ss}^{(s-1)}}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} v_s^2} dv_s$$

Avem însă relațiile cunoscute :

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2} v_s^2} dv_s = 1,\ |a_{kh}| = a_{11}\, a_{22}^{(1)}\, a_{33}^{(2)} \ldots a_{ss}^{(s-1)}$$

de unde rezultă relația care trebuia demonstrată (7_3) :

$$h_n = \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} f(x_1, x_2, \ldots, x_s) \prod_{k=1}^{s} dx_k = 1$$

Observații. 1°. Să notăm prin $\| a^{kh} \|$ inversa matricei $\| a_{kh} \|$ a formei pătratice $\sum_{k=1}^{s} \sum_{h=1}^{s} a_{kh} (x_k - m_k)(x_h - m_h)$, adică :

$$\| a^{kh} \| = \| a_{kh} \|^{-1}$$

Dacă în $\| a^{kh} \|$ suprimăm prima linie și prima coloană obținem o matrice pătrată de ordinul (s—1) Determinantul inversei acestei matrice, după cum am văzut, este tocmai raportul $\frac{\sqrt{|a_{kh}|}}{\sqrt{a_{11}}}$.

După a doua integrare, va trebui să suprimăm în $\| a^{kh} \|$ primele două linii și două coloane și să luăm inversa matricei obținute. Determinantul acestei matrici este raportul:

$$\frac{\sqrt{|a_{kh}|}}{\sqrt{a_{11}} \cdot \sqrt{a_{22}^{(1)}}}.$$

După (s—1) integrări obținem:

$$\frac{1}{\sqrt{2\pi}} \sqrt{a}\ e^{-\frac{1}{2} a\, \xi_s^2}$$

unde a este funcție de elementele inițiale ale matricei prin relațiile succesive (7_5), $(7_5')$, $(7_5'')$, și este totodată matricea obținută, din $\| a^{kh} \|$, suprimînd primele (s-1) linii și coloane, formată, deci, dintr-un singur element a^{ss}; inversa acestei matrici este $a = \frac{1}{a_{ss}}$ și este egală cu raportul:

$$\frac{\sqrt{|a_{kh}|}}{\sqrt{a_{11}} \cdot \sqrt{a_{22}^{(1)}} \ldots \sqrt{a_{s-1,\, s-1}^{(s-2)}}}.$$

2. *Distribuția marginală* în raport cu primele r variabile $x_1, x_2, \ldots x_r$, se obține integrînd densitatea (7) în raport cu ultimele s—r variabile $x_{r+1}, \ldots x_s$ între $-\infty$ și $+\infty$. Rezultatul este de forma:

$$\left(\frac{1}{2}\right)^{\frac{r}{2}} \sqrt{|\bar{a}_{kh}|}\ e^{-\frac{1}{2} \sum_{k=1}^{r} \sum_{h=1}^{r} \bar{a}_{kh} (x_k - m_k)(x_h - m_h)}$$

unde k și h iau valorile 1, 2,..., r. Coeficienții $\bar{a}_{kh}$ se obțin din $\| a^{kh} \|$ suprimînd de astă dată ultimele (s—r) linii și coloane și luînd inversa matricei obținute, adică:

$$\| \bar{a}_{kh} \| = \| a^{kh} \|^{-1},\ k\ h = 1, 2, \ldots, r$$

3°. *Distribuția condiționată* pentru primele r variabile $x, x_2, \ldots x_r$ va fi:

$$g_{x_{r+1}, \ldots x_s}(x_1, x_2, \ldots, x_r) = \frac{f(x_1, x_2, \ldots, x_s)}{g(x_{r+1}, \ldots, x_s)}$$

unde $g(x_{r+1}, \ldots, x_s)$ este densitatea marginală pentru $x_{r+1}, \ldots, x_s$, adică:

$$g(x_{r+1}, \ldots, x_s) = \left(\frac{1}{2\pi}\right)^{\frac{s-r}{2}} \sqrt{|\bar{a}_{kh}|}\ e^{-\frac{1}{2} \sum_{k=r+1}^{s} \sum_{h=r+1}^{s} \bar{a}_{kh} (x_k - m_k)(x_h - m_h)}$$

iar $\| \bar{a}_{kh} \| = \| \bar{a}^{kh} \|^{-1}$

Scanned with OKEN Scanner

§ 3. Distribuții de funcții de variabile aleatorii

În capitolul VI, § 2 am definit o funcție :

$$u = u(x) \tag{8}$$

de variabilă aleatorie x cu densitatea f(x) și am calculat densitatea g(u) a variabilei u în cazul particular cînd funcția u(x) este monoton crescătoare. Am găsit atunci densitatea :

$$g(u) = f[x(u)] \left| \frac{du}{dx} \right| \tag{8_1}$$

Acum vom generaliza această formulă pentru o funcție u(x) nemonotonă.

Să notăm x=x(u) funcția inversă a funcției (8) ; în general, funcția x(u) este o funcție multiformă deci la o valoare dată u, corespund mai multe valori pentru x, de exemplu :

$$x_1(u), \quad x_2(u), \ \ldots \ x_r(u) \tag{8_2}$$

Să considerăm un interval Δu în jurul lui u pe axa o u ; acestuia îi corespund intervalele Δx_1, Δx_2... Δx_r respectiv în jurul

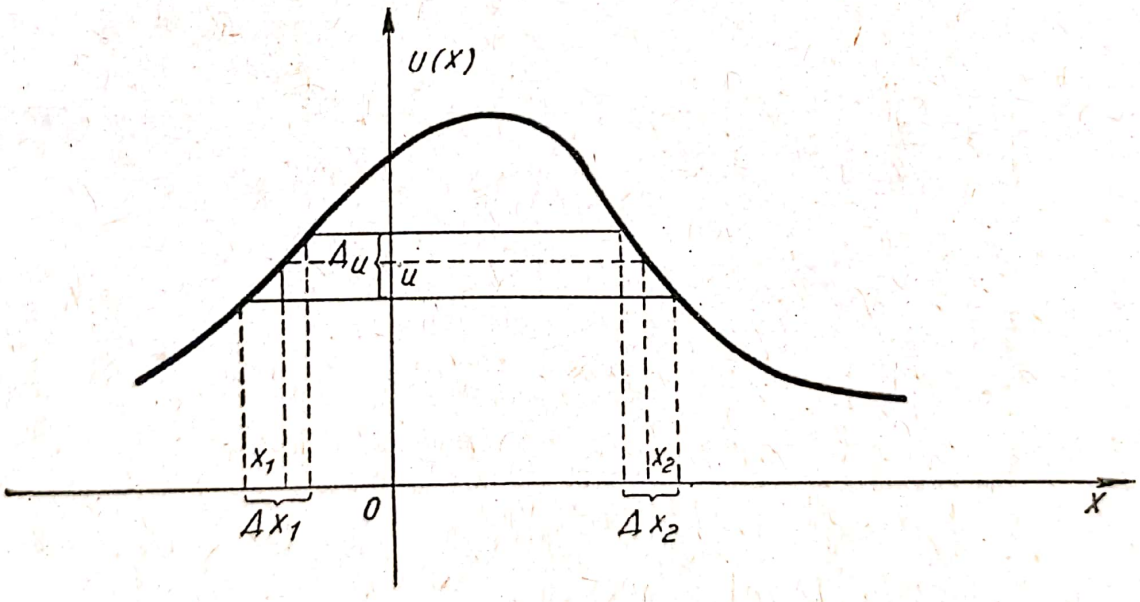

Fig. 44

punctelor $x_1, x_2, \ldots, x_r$ de pe axa ox (În graficul alăturat r=2). Notînd cu P (u ∈ Δu) probabilitatea ca u să se găsească pe Δu, după teorema de adunare a probabilităților, aceasta este egală

cu suma probabilităților ca x să se găsească sau pe Δx_1, sau pe Δx_2,... sau pe Δx_r, indiferent pe care, deci :

$$P(u \in \Delta u) = \sum_{k=1}^{r} P(x_k \in \Delta x_k) \qquad (8_3)$$

Avem însă relația :

$$P(u \in \Delta u) = \int_{\Delta u} g(u)\, du$$

unde g (u) este densitatea lui u, iar integrala este luată pe segmentul Δu. După formula mediei rezultă :

$$P(u \in \Delta u) = \int_{\Delta u} g(u)\, du = g(u')\, \Delta u, \quad (u' \in \Delta u)$$

și în mod analog :

$$P(x_k \in \Delta x_k) = f(x_k')\, \Delta x_k, \quad (x_k' \in \Delta x_k)$$

unde $f(x_k)$ este densitatea variabilei x_k.

Relația (8_3) devine :

$$g(u')\, \Delta u = \sum_{k=1}^{r} f(x_k')\, \Delta x_k$$

De unde, împărțind prin Δu și trecînd la limită pentru $\Delta u \to 0$, deci și $\Delta x_k \to 0$, așa încît,

$$u' \to u$$

$$x_k' \to x_k, \text{ pentru orice } k = 1, 2, \ldots, r$$

Vom obține :

$$g(u) = \sum_{k=1}^{r} f(x_k) \frac{dx_k}{du}$$

Dacă unele derivate $\frac{dx_k}{du}$ sînt negative, relația (8_3), în care toți termenii din partea dreaptă sînt pozitivi, ne arată că trebuie să schimbăm semnul acestor derivate. Atunci vom avea formula generală :

$$g(u) = \sum_{k=1}^{r} f[x_k(u)] \left| \frac{dx_k(u)}{du} \right| \qquad (8_4)$$

Scanned with OKEN Scanner

Exemple: 1°. Să se calculeze distribuția variabilei $u = \sqrt{x^2 + 3} - 1$ știind că densitatea variabilei x este:

$$f(x) = \frac{1}{12}(x + 2) \qquad -1 \leqslant x \leqslant 3$$

Graficul funcției u (x) este cel redat mai jos:

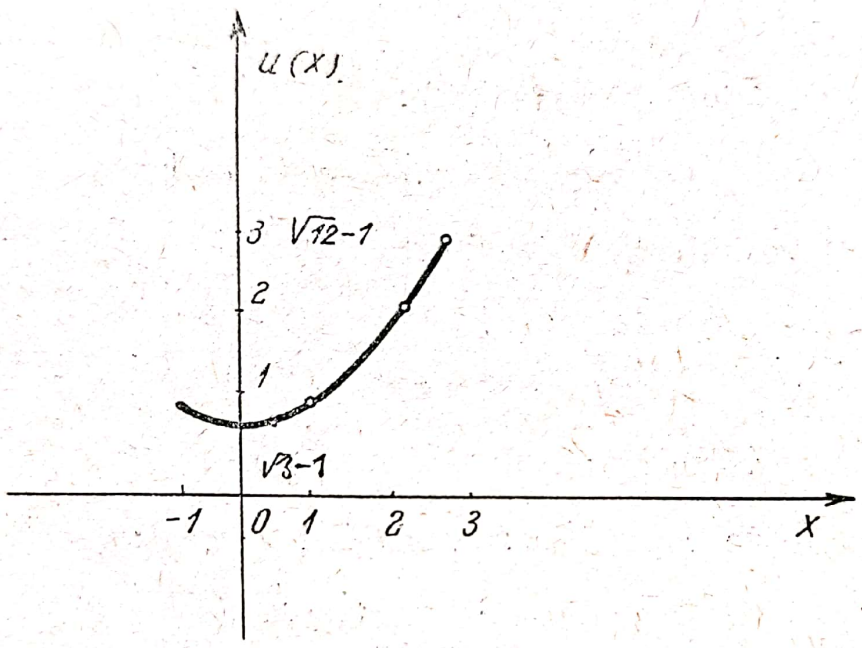

Fig. 45.

Intervalul de variație al lui u este de la $\sqrt{3} - 1$, pentru $x = 0$, pînă la $\sqrt{12} - 1$, pentru $x = 3$.

Dacă $u < 1$, la o valoare u corespunde funcția inversă

$$x_1(u) = -\sqrt{u^2 + 2u - 2}, \qquad (x < 0)$$

$$x_2(u) = +\sqrt{u^2 + 2u - 2}, \qquad (x > 0)$$

de unde rezultă densitatea:

$$g(u) = \frac{1}{12}\left[-\sqrt{u^2 + 2u - 2} + 2\right]\frac{u + 1}{\sqrt{u^2 + 2u - 2}} +$$

$$+ \frac{1}{12}\left[\sqrt{u^2 + 2u - 2} + 2\right]\frac{u + 1}{\sqrt{u^2 + 2u - 2}} = \frac{1}{3} \cdot \frac{u + 1}{\sqrt{u^2 + 2u - 2}}$$

Scanned with OKEN Scanner

Pentru $u > 1$ avem funcția inversă : $x(u) = \sqrt{u^2 + 2u - 2}$ deci densitatea :

$$g(u) = \frac{1}{12}\left[\sqrt{u^2 + 2u - 2} + 2\right] \frac{u + 1}{\sqrt{u^2 + 2u - 2}} \qquad (x > 0)$$

2°. Să se calculeze densitatea funcției $u(x) = x^2$ știind că densitatea variabilei x este $f(x)$.

Avem funcția inversă :

$$x_1(u) = -\sqrt{u}, \qquad (x < 0)$$
$$x_2(u) = +\sqrt{u}, \qquad (x > 0)$$

De unde aplicînd formula (8_4) :

$$g(u) = \frac{1}{2\sqrt{u}}\left[f(-\sqrt{u}) + f(+\sqrt{u})\right] \qquad (8_5)$$

3°. Să se calculeze densitatea funcției

$$u(x) = \sqrt{x} \qquad (x > 0)$$

Funcția inversă este $x(u) = u^2$, deci vom găsi densitatea :

$$g(u) = 2\,u\,f(u^2) \qquad (8_6)$$

Observație. Problemele de mai sus se pot rezolva și în modul următor :

Funcția de repartiție a variabilei u este :

$$G(u) = \int_{u(x) < u} f(x)\,dx$$

unde integrala este luată după toate valorile de pe axa ox pentru care $u(x) < u$.

Așa de pildă în exemplul 2° găsim :

$$G(u) = \int_{u(x) < u} f(x)\,dx = \int_{x^2 < u} f(x)\,dx$$

Dar inegalitatea $x^2 < u$ este echivalentă cu : $-\sqrt{u} < x < +\sqrt{u}$, așa încît, vom putea scrie : $G(u) = F(+\sqrt{u}) - F(-\sqrt{u})$

Densitatea $g(u)$ se obține prin derivare :

$$g(u) = G'(u) = F'(u) \cdot \frac{1}{2\sqrt{u}} + F'(-\sqrt{u}) \cdot \frac{1}{2\sqrt{u}} =$$

$$= \frac{1}{2\sqrt{u}}\left[f(-\sqrt{u}) + f(+\sqrt{u})\right]$$

adică formula (8_4).

Distribuția funcțiilor de variabile aleatorii. Fie $x_1, x_2, \ldots x_s$, s variabile aleatorii, $f(x_1, x_2, \ldots, x_s)$ densitatea asociată lor și funcția :

$$u = u(x_1, x_2, \ldots, x_s) \tag{9}$$

Densitatea variabilei u se obține ca densitatea unei distribuții marginale, în modul următor :

Rezolvăm relația (9) în raport cu una din variabile, x_1 de pildă, și obținem în general un sistem de r soluții, fie

$$x_{1k} = x_{1k}(u, x_2, \ldots x_s), \qquad 1 \leqslant k \leqslant r \tag{9_1}$$

Raționînd ca și în cazul unidimensional obținem densitatea pentru u și variabilele $x_2, x_3, \ldots, x_s$;

$$g(u, x_2, x_3, \ldots x_s) = \sum_{k=1}^{r} f[x_{1k}(u, x_2, \ldots, x_s), x_2, \ldots x_s] \left| \frac{\partial x_{1k}}{\partial u} \right| \tag{9_2}$$

Densitatea lui u se obține ca orice densitate marginală, integrînd funcția $g(u, x_2, \ldots x_s)$ în raport cu variabilele $x_2, x_3, \ldots, x_s$ între limitele $-\infty$ și $+\infty$.

Exemplu. Fie distribuția : $f(x_1, x_2) = \frac{1}{8}(1 - x_1 - x_2)$, pentru $0 < x_1 < 2$; $2 < x_2 < 4$ și funcția : $u(x_1, x_2) = 1 - x_1 + x_2^2$.

Ne propunem să calculăm densitatea funcției $u(x_1, x_2)$.

Rezolvînd în raport cu x_1 găsim : $x_1(u, x_2) = 1 - u + x_2^2$

Densitatea asociată acestor variabile se calculează cu formula (9_2)

$$g(u, x_2) = f[x_1(u, x_2), x_2] \left| \frac{\partial x_1}{\partial u} \right| = \frac{1}{8}(5 + u - x_2^2 - x_2)$$

Densitatea marginală în raport cu u este :

$$g(u) = \int_{-\infty}^{+\infty} g(u, x_2)\, dx_2 = \int_2^4 \frac{1}{8}(5 + u - x_2^2 - x_2)\, dx_2 = \frac{1}{3}(9u - 29)$$

2) Dacă avem p funcții de s variabile aleatorii, $p \leqslant s$:

$$\begin{aligned} u_1 &= u_1(x_1, x_2, \ldots, x_s) \\ u_2 &= u_2(x_1, x_2, \ldots, x_s) \\ &\cdot \\ &\cdot \\ &\cdot \\ u_p &= u_p(x_1 x_2, \ldots, x_s) \end{aligned} \tag{9_3}$$

Scanned with OKEN Scanner

rezolvăm acest sistem în raport cu p variabile; fie, de pildă acestea : $x_1, x_2, \ldots x_p$:

$$\begin{aligned} x_1 &= x_1 (u_1, u_2, \ldots u_p, x_{p+1} \ldots x_s) \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ x_p &= x_p (u_1, u_2, \ldots u_p, x_{p+1}, \ldots x_s) \end{aligned} \tag{9_4}$$

Densitatea asociată variabilelor $u_1, \ldots u_p, x_{p+1}, \ldots x_s$ se obține ca și mai înainte :

$$g (u_1, u_2, \ldots, u_p, x_{p+1}, \ldots x_s) =$$

$$= \Sigma f [x_1, (u_1 u_2, \ldots, u_p, x_{p+1} \ldots, x_s) \ldots x_p (u_1, \ldots u_p, x_{p+1} \ldots x_s] \left| \frac{\partial x_k}{\partial u_h} \right|^* \tag{9_5}$$

unde suma se extinde asupra tuturor soluțiilor (sau sistemelor de soluții) din (9_4) și unde prin $\left| \frac{\partial x_k}{\partial u_h} \right|^*$ am însemnat următorul determinant funcțional :

$$\left| \frac{\partial x_k}{\partial u_h} \right| = \frac{D (x_1, x_2, \ldots x_p)}{D (u_1, u_2, \ldots, u_p)} = \begin{vmatrix} \frac{\partial x_1}{\partial u_1} & \cdots\cdots & \frac{\partial x_1}{\partial u_p} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \frac{\partial x_p}{\partial u_1} & \cdots\cdots & \frac{\partial x_p}{\partial u_p} \end{vmatrix} \tag{9_6}$$

luat în valoare absolută.
Formula (9_5) este aceeași cu formula schimbării de variabile într-o integrală multiplă. De exemplu, dacă în integrala dublă

$$I = \iint_D f (x, y)\, dx\, dy$$

facem schimbarea de variabile

$$x = x (u, v), \qquad y = y (u, v)$$

$$I = \iint_D f [x (u, v), \quad y (u, v)] \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}^+ du\, dv$$

Altă metodă. Determinarea densității distribuțiilor de funcții de variabile aleatorii se poate face folosind și funcția generatoare

de momente sau funcția caracteristică. Așa de exemplu pentru funcția (9) putem scrie :

$$m(t) = M\left[e^{t\,u(x_1, x_2, \ldots, x_s)}\right] =$$

$$= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, x_2, \ldots x_s)\, e^{t\,u(x_1, x_2, \ldots, x_s)}\, dx_1\, dx_2 \ldots dx_s \tag{10}$$

Dacă m (t) este funcția generatoare a unei distribuții cunoscute, evident u va avea această distribuție, în virtutea unei teoreme dată în capitolul precedent.

Distribuția lui χ^2. Se dau variabilele independente $x_1, x_2 \ldots, x_s$ normal distribuite, avînd toate mediile egale cu zero și dispersiile egale cu 1, adică pentru orice $k = 1, 2, \ldots, s$ avem densitatea

$$f(x_k) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_k^2} \tag{11}$$

Să calculăm densitatea funcției :

$$\chi^2(x_1, x_2, \ldots, x_s) = x_1^2 + x_2^2 + \ldots + x_s^2 \tag{11_1}$$

Am văzut (8_5) că variabilele x_k^2 au densitatea :

$$\frac{1}{2\sqrt{x_k}}\left[f(-\sqrt{x_k}) + f(+\sqrt{x_k})\right] = x_k^{-\frac{1}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_k}$$

Funcția generatoare a variabilei x_k va fi :

$$m_k(t) = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} x_k^{-\frac{1}{2}} e^{-\frac{1}{2}x_k} e^{tx_k}\, dx_k = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} x_k^{-\frac{1}{2}} e^{-\frac{(1-2t)x_k}{2}}\, dx_k.$$

După definiția funcției Gama putem scrie :

$$m_k(t) = \frac{1}{\sqrt{2\pi}} \cdot \frac{\Gamma\left(\frac{1}{2}\right)}{\left(\frac{1-2t}{2}\right)^{1/2}} = \frac{1}{(1-2t)^{1/2}}$$

pentru orice $k = 1, 2, \ldots s$.

Funcția generatoare a sumei $\sum_{k=1}^{s} x_k^2$ este egală cu produsul funcțiilor generatoare ale termenilor, deci :

$$m(t) = \frac{1}{(1-2t)^{s/2}} \tag{11_2}$$

Scanned with OKEN Scanner

Această funcție este un caz particular al funcției generatoare a distribuției Gama și anume pentru :

$$a + 1 = \frac{s}{2}, \quad b = 2 \tag{11_3}$$

Întrucît densitatea funcției Gama este :

$$f(x) = \frac{1}{\Gamma(a+1)\, b^{a+1}} x^a \, e^{-\frac{x}{b}}, \quad x > 0$$

urmează că densitatea distribuției $\xi = \chi^2$ va fi :

$$f(\xi) = \frac{1}{\Gamma\left(\frac{s}{2}\right) 2^{\frac{s}{2}}} \cdot \xi^{\frac{s}{2}-1} \, e^{-\frac{\xi}{2}} \tag{11_4}$$

Din cauza parametrului s care intră în această formulă și din cauză că χ^2 este suma a s pătrate, se spune că densitatea χ^2 are *s grade de libertate*. Pentru această distribuție statisticianul englez K. Pearson a construit tabele speciale de calcul.

Distribuția Z a lui R. A. Fischer.

Să presupunem că variabilele independente u și v au distribuții χ^2 respectiv cu r și s grade de libertate. După (11_4) ele au densitățile :

$$\varphi(u) = \frac{1}{\Gamma\left(\frac{r}{2}\right) 2^{\frac{r}{2}}} \cdot u^{\frac{r}{2}-1} \, e^{-\frac{u}{2}}$$

$$\psi(v) = \frac{1}{\Gamma\left(\frac{s}{2}\right) 2^{\frac{s}{2}}} \cdot v^{\frac{s}{2}-1} \, e^{-\frac{v}{2}}.$$

Întrucît u și v sînt independente distribuția lor va avea densitatea $f(u, v) = \varphi(u) \cdot \psi(v)$, adică :

$$f(u, v) = \frac{1}{\Gamma\left(\frac{r}{2}\right)\Gamma\left(\frac{s}{2}\right) 2^{\frac{r+s}{2}}} u^{\frac{r}{2}-1} v^{\frac{s}{2}-1} e^{-\frac{1}{2}(u+v)} \tag{12}$$

Scanned with OKEN Scanner

Ne propunem să calculăm densitatea variabilei :

$$Z = Z(u, v) = \frac{\left(\frac{u}{r}\right)}{\left(\frac{v}{s}\right)} = \frac{s\,u}{r\,v} \qquad (12_1)$$

Problema ne este cunoscută : eliminăm una din variabile, u de pildă și obținem distribuția g(v,Z) a variabilelor v și Z ; integrînd în raport cu v de la 0 la ∞ găsim densitatea căutată g(Z). Din (12_1) avem $u = \frac{r\,v\,Z}{s}$; introducînd în (12) se găsește :

$$g(v, Z) = \frac{1}{\Gamma\left(\frac{r}{2}\right) \cdot \Gamma\left(\frac{s}{2}\right) 2^{\frac{r+s}{2}}} \left(\frac{r}{s}\right)^{\frac{r}{2}-1} Z^{\frac{r}{2}-1}\, v^{\frac{r+s}{2}-2}\, e^{-\frac{r}{2}\left(1+\frac{rz}{s}\right)} \left|\frac{\partial u}{\partial z}\right|$$

Înlocuim $\frac{\partial u}{\partial z} = \frac{rv}{s}$ și apoi integrăm în raport cu v de la 0 la ∞, obținem densitatea lui Z :

$$g(Z) = \frac{\left(\frac{r}{s}\right)^{\frac{r}{2}} Z^{\frac{r}{2}-1}}{\Gamma\left(\frac{r}{2}\right) \cdot \Gamma\left(\frac{s}{2}\right) 2^{\frac{r+s}{2}}} \int_0^{\infty} v^{\frac{r+s}{2}-1}\, e^{-\left(\frac{s+rz}{2s}\right)v}\, dv =$$

$$= \frac{\left(\frac{r}{s}\right)^{\frac{r}{2}} Z^{\frac{r}{2}-1}}{\Gamma\left(\frac{r}{2}\right) \cdot \Gamma\left(\frac{s}{2}\right) 2^{\frac{r+s}{2}}} \cdot \frac{\Gamma\left(\frac{r+s}{2}\right)}{\left(\frac{rz+s}{2s}\right)^{\frac{r+s}{2}}}$$

sau încă :

$$g(Z) = \frac{\Gamma\left(\frac{r+s}{2}\right) \cdot r^{\frac{r}{2}} \cdot s^{\frac{s}{2}}}{\Gamma\left(\frac{r}{2}\right) \cdot \Gamma\left(\frac{s}{2}\right)} \cdot \frac{Z^{\frac{r-2}{2}}}{(rz+s)^{\frac{r+s}{2}}} \qquad (12_2)$$

Scanned with OKEN Scanner

Introducînd funcția Beta (relația 6_2, capitolul V) densitatea de mai sus se poate scrie :

$$g(Z) = \frac{r^{\frac{r}{2}} s^{\frac{s}{2}}}{\beta\left(\frac{r}{2}, \frac{s}{2}\right)} \cdot \frac{Z^{\frac{r-2}{2}}}{(rZ+s)^{\frac{r+s}{2}}} \qquad (12'_2)$$

Distribuția variabilei Z se spune că are r și s grade de libertate.

Distribuția t a lui Student. Variabilele u și v sînt independente ; u are o distribuție normală, iar v o distribuție egală cu rădăcina pătrată a variabilei χ^2, adică $v = \chi$ cu r grade de libertate. Vom avea densitățile :

$$f(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$$

$$\varphi(v) = \frac{1}{\Gamma\left(\frac{r}{2}\right) 2^{\frac{r}{2}}} \cdot v^{\frac{r}{2}-1} e^{-\frac{1}{2}v}$$

Densitatea distribuției bidimensionale în u și v este :

$$f(u, v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} \cdot \frac{1}{\Gamma\left(\frac{r}{2}\right) 2^{\frac{r}{2}}} \cdot v^{\frac{r}{2}-1} e^{-\frac{1}{2}v}$$

Să găsim acum densitatea funcției :

$$t = \frac{u}{\sqrt{\frac{v}{r}}} \qquad (13)$$

numită distribuția lui Student[1].

[1] *W. S. Gosett* a scris sub pseudonimul Student.

Eliminăm pe $u = t\sqrt{\frac{v}{r}}$ și găsim :

$$g(t, v) = \frac{1}{\sqrt{2\pi} \cdot \Gamma\left(\frac{r}{2}\right) \cdot 2^{\frac{r}{2}}} e^{-\frac{1}{2}t^2 \frac{v}{r}} v^{\frac{r}{2}-1} e^{-\frac{1}{2}v} \cdot \frac{v}{\sqrt{r}}^{\frac{1}{2}}$$

deoarece : $\left|\frac{\partial u}{\partial t}\right| = \sqrt{\frac{v}{r}} = \frac{v^{\frac{1}{2}}}{\sqrt{r}}$.

Integrînd în raport cu v de la 0 la $+\infty$ obținem densitatea lui t :

$$g(t) = \frac{1}{\sqrt{\pi r} \cdot \Gamma\left(\frac{r}{2}\right) 2^{\frac{r+1}{2}}} \int_0^{\infty} v^{\frac{r-1}{2}} e^{-\frac{1}{2}\left(1+\frac{t^2}{2}\right)v} dv =$$

$$= \frac{1}{\sqrt{\pi r} \cdot \Gamma\left(\frac{r}{2}\right) 2^{\frac{r+1}{2}}} \cdot \frac{\Gamma\left(\frac{r+1}{2}\right)}{\left(1 + \frac{t^2}{r}\right)^{\frac{r+1}{2}}} \cdot 2^{\frac{r+1}{2}}$$

Așadar, distribuția lui t va fi :

$$g(t) = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi r} \cdot \Gamma\left(\frac{r}{2}\right)} \cdot \frac{1}{\left(1 + \frac{t^2}{r}\right)^{\frac{r+1}{2}}}, \quad -\infty < t < +\infty \qquad (13_1)$$

adică o familie depinzînd de un singur parametru r.

Probleme

1. Se dă distribuția :

$$f(x, y) = \frac{\sqrt{5}}{6\pi} e^{-\frac{1}{6}[4(x+1)^2 - 2(x+1)(y-2) + (y-2)^2]}, \quad -\infty < x, y < +\infty$$

Să se determine :
a) Valorile medii, dispersiile și covariațele variabilelor
b) Funcția generatoare de momente

Scanned with OKEN Scanner

2. O distribuție normală cu trei variabile are la exponent forma pătratică : $f(x, y, z) = 2x^2 + y^2 + 4z^2 - xy - 2xz$
Să se determine :

a) Dispersiile și covariațele variabilelor x, y, z.

b) Densitatea marginală în raport cu z.

c) Densitatea condiționată a variabilelor x și x și z

3. Se consideră o distribuție s-dimensională cu densitatea $f(x_1, x_2 \ldots x_s)$ pentru care se definesc :

$$m_i = M(x_k), \quad 1 \quad i \quad o, \text{ valorile medii}$$

$$\sigma_{ii} = M(x_i - m_i)^2, \text{ dispersiile}$$

$$\sigma_{ij} = M(x_i - m_i)(x_y - m_y), \text{ covariațele,}$$

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii} \cdot \sigma_{jj}}}, \text{ corelațiile.}$$

Să se calculeze :

a) Valoarea medie și dispersia funcției $y = a_1 x_1 + a_2 x_2 + \ldots a_s x_s$

b) Corelația între funcțiile $y = \Sigma a_i x_i$ și $z = \Sigma b_i x_i$

4. Densitatea condiționată în raport cu primele r variabile din problema precedentă este :

$$f(x_1, x_2 \ldots x_r / x_{rn} \ldots x_s) = \frac{f(x_1, x_2 \ldots x_s)}{g(x_{rn}, \ldots x_s)}$$

unde $g(x_{rn} \ldots x_s)$ este densitatea marginală în raport cu $x_{rn} \ldots x_s$.

Distribuția condiționată are valori medii, funcții de $x_{rn} \ldots x_s$ pe care le notăm prin : $m_i (x_{rn}, \ldots x_s)$ $u = 1, 2, \ldots r$ și care se numesc *funcții de regresiune.*

Să se demonstreze că valoarea probabilă a funcției de regresiune $m_i (x_{rn}, \ldots x_s)$ este valoarea medie a lui x_i în distribuția necondiționată.

5. Notînd prin $\sigma_{ij} (x_{rn}, \ldots x_s)$ covariațele variabilelor din problema 4, (i, j = 1, 2 ... r) să se arate că acestea sînt constante pentru distribuția normală bidimensională *(generalizare).*

6. Dacă se notează : $M(\sigma_{ij} (x_{rn}, \ldots x_s)) = \sigma_{ij \,.\, (r+1) \ldots s}$ acestea se numesc covariațele în raport cu funcțiile de regresiune, iar coeficienții :

$$\rho_{ij \,.\, (rn) \ldots s} = \frac{\sigma_{ij \,.\, (rn) \ldots s}}{\sqrt{\sigma_{ij \,.\, (rn) \ldots s} \, \sigma_{ij \,(rn) \ldots s}}}$$

Scanned with OKEN Scanner

coeficienți parțiali de corelație ai distribuției condiționate.

a) Să se calculeze $\sigma_{11.2}$ pentru distribuția normală bidimensională;

b) Să se arate că pentru distribuția normală tridimensională avem:

$$\rho\, 12.3 = \frac{\rho_{12} - \rho_{13}\,\rho_{23}}{\sqrt{(1-\rho_{13}^2)\,(1-\rho_{23}^2)}}$$

7. Se notează $\xi_{1\,.23\,\ldots\,s}$ abaterea lui x_1 față de funcția sa de regresiune în distribuția condiționată a lui x_1. Pentru o distribuție normală cu trei variabile x_1, x_2, x_3 să se arate că x_1, $x_{2.1}$ $x_{3.21}$ au distribuții normale independente.

8. Într-o distribuție normală s-dimensională se consideră distribuția condiționată a lui x_1 și funcția de regresiune y. Corelația dintre x_1 și y se numește coeficient de corelație multiplă al lui x_1 față de y și se notează $R_{1\,.23\,\ldots\,s}$. Să se arate că pentru o distribuție normală tridimensională avem: $\sigma_{11.23} = \sigma_{11}\,(1 - R^2_{1.23})$ și în general: $\sigma_{11.23\ldots s} = \sigma_{11}\,(1 - R^2_{1.23\ldots s})$

9. Să se determine densitatea variabilei: $Z = \frac{1}{2}\,\ln\,\frac{su}{rv}$ unde u și v au distribuții χ^2 respectiv cu r și s grade de libertate.

Capitolul VIII

LEGEA NUMERELOR MARI. TEOREME DE CONVERGENȚĂ

§1. Convergența în probabilitate

Să considerăm șirul de n variabile aleatorii

$$x_1, x_2, \ldots x_n \tag{1}$$

definite pe același cîmp de probabilitate [E, k, P].

D e f i n i ț i e . Șirul de variabile aleatorii (1) *converge sau tinde în probabilitate* către numărul constant a, dacă pentru orice ε pozitiv avem relația :

$$\lim_{n \to \infty} P(|x_n - a| < \varepsilon) \to 1 \tag{2}$$

Așadar, dacă n este suficient de mare, variabilele x_n se abat de la numărul a cu mai puțin decît ε, aproape sigur.

Sub o formă echivalentă relația (2) se mai poate scrie :

$$\lim_{n \to \infty} P(|x_n - a| > \varepsilon) \to 0 \tag{2'}$$

T e o r e m ă. Condiția suficientă ca șirul (1) să tindă în probabilitate către numărul a, este

$$\lim_{n \to \infty} M(x_n - a)^2 \to 0 \tag{3}$$

Demonstrație. Prin definiție, avem :

$$M(x_n - a)^2 = \int_{-\infty}^{+\infty} (x_n - a)^2 f_n(x)\, dx \tag{3_1}$$

unde $f_n(x)$ reprezintă densitatea de probabilitate a variabilei x_n. Integrala din partea dreaptă se poate scrie :

$$\int_{-\infty}^{+\infty} = \int_{-\infty}^{a-\varepsilon} + \int_{a-\varepsilon}^{a+\varepsilon} + \int_{a+\varepsilon}^{\infty} \qquad (3_2)$$

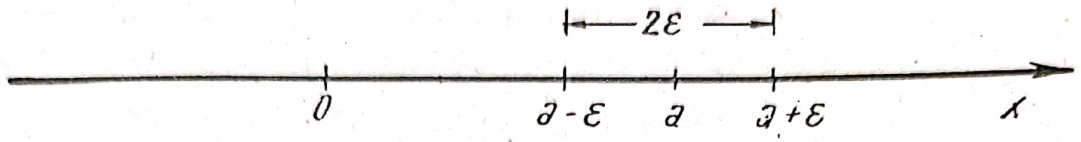

Fig. 46

În intervalul $(a - \varepsilon,\ a + \varepsilon)$ avem : $|x_n - a| < \varepsilon$, iar în afara acestui interval $|x_n - a| > \varepsilon$
Dacă suprimăm în partea dreaptă a relației (3_2) termenul pozitiv

$$\int_{a-\varepsilon}^{a+\varepsilon} (x_n - a)^2 f_n(x)\, dx$$

iar în celelalte două integrale înlocuim $(x_n - a)^2$ prin numărul mai mic ε^2 (căci în aceste integrale x_n ia valori în afara intervalului de la $a - \varepsilon$ la $a + \varepsilon$) evident, inegalitatea se poate scrie :

$$\int_{-\infty}^{+\infty} (x_n - a)^2 f_n(x)\, dx \geqslant \int_{-\infty}^{a-\varepsilon} \varepsilon^2 f_n(x)\, dx + \int_{a+\varepsilon}^{+\infty} \varepsilon^2 f_n(x)\, dx$$

De unde rezultă, ținînd seama de relația (3_1) :

$$\frac{M(x_n - a)^2}{\varepsilon^2} \geqslant \int_{-\infty}^{a-\varepsilon} f_n(x)\, dx + \int_{a+\varepsilon}^{+\infty} f_n(x)\, dx$$

Suma integralelor reprezintă însă probabilitatea ca x_n să ia valori în afara intervalului $(a - \varepsilon,\ a + \varepsilon)$, adică probabilitatea ca $|x_n - a| > \varepsilon$, așa încît

$$\frac{M(x_n - a)^2}{\varepsilon^2} \geqslant P(|x_n - a| > \varepsilon) \qquad (4)$$

Putem deci scrie :

$$0 \leqslant P(|x_n - a| > \varepsilon) = P(|x_n - a|^2 > \varepsilon^2) \leqslant \frac{M(x_n - a)^2}{\varepsilon^2}$$

În baza relațieiei (3) obținem deci limita

$$\lim_{n \to \infty} P(|x_n - a| > \varepsilon) \to 0$$

adică relația (21) care arată că șirul (1) converge în probabilitate către a.

Observație. Deoarece evenimentele $|x_n - a| > \varepsilon$ și $|x_n - a| < \varepsilon$ sînt evenimente contrare, avem relația: $P(|x_n - a| > \varepsilon) = 1 - P(|x_n - a| < \varepsilon)$

Inegalitatea (4) devine:

$$P(|x_n - a| < \varepsilon) \geqslant 1 - \frac{M(x_n - a)^2}{\varepsilon^2} \qquad (4')$$

În cazul particular, cînd $a = M(x)$, avem:

$$M(x_n - a)^2 = M[x_n - M(x_n)]^2 = D^2(x_n) = \sigma_n^2$$

Dacă suprimăm indicele n, ajungem la concluzia că pentru orice variabilă aleatorie x cu media M (x) și dispersia finită σ^2 subzistă inegalitatea următoare:

$$P(|x - M(x)| < \varepsilon) \geqslant 1 - \frac{\sigma^2}{\varepsilon^2} \qquad (4'')$$

Această inegalitate stabilită prima oară de Bienaymé a fost pusă în adevărata sa lumină de Cebîșev.

Inegalitatea $|x - M(x)| < \varepsilon$ este echivalentă cu dubla inegalitate $M(x) - \varepsilon < x < M(x) + \varepsilon$ în consecință, inegalitatea (4'') a lui Cebîșev dă o limită inferioară a probabilității ca variabila aleatorie x să ia o valoare în intervalul $[M(x) - \varepsilon, M(x) + \varepsilon)]$.

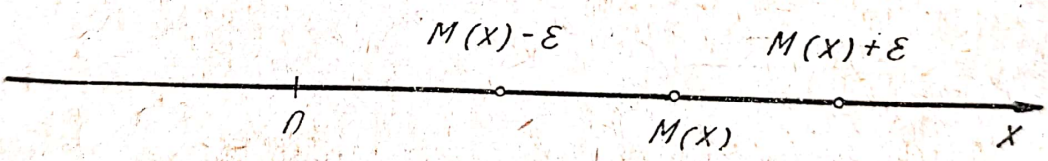

Fig. 47

De obicei, ε se exprimă în unități egale cu σ, adică se ia $\varepsilon = k\sigma$; atunci relația (4'') se scrie:

$$P(|x - M(x)| < k\sigma) \geqslant 1 - \frac{1}{k^2} \qquad (4_1''')$$

Pentru $k = 3$, de pildă, obținem: $P(|x - M(x)| < 3\sigma) \geqslant \frac{8}{9} \approx 0{,}888\ldots$

Să lămurim sensul unui asemenea rezultat: Se face un număr de n probe independente și în fiecare probă evenimentul A

are aceeași probabilitate p de a se realiza. Am găsit în capitolele precedente valorile : $M(x) = np$, $D^2(x) = \sigma^2 = npq$.

Fie $n = 100$, $p = q = \frac{1}{2}$, deci $M(x) = 50$, $\sigma = 5$. Intervalul $[M(x) - 3\sigma, M(x) + 3\sigma] = [35, 65]$; inegalitatea lui Cebîșev afirmă că dacă se face un număr mare de serii de cîte 100 probe, 1 000 de pildă, în cel puțin 888 din ele, evenimentul A se va realiza între 35 și 65 ori.

Teorema lui Cebîșev. Dacă un șir de variabile aleatorii

$$x_1, x_2, \ldots, x_n \tag{5}$$

satisfac condițiile :

1° Sînt independente între ele ;

2° Au aceeași valoare medie, s-o notăm cu m :

$$M(x_1) = M(x_2) = \ldots = M(x_n) = m \tag{5_1}$$

3° Au dispersiile egal mărginite, adică există un număr constant L așa fel că pentru orice $k = 1, 2, \ldots n$ avem :

$$D^2(x_k) < L \tag{5_2}$$

Atunci media aritmetică $\bar{x} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n)$ a variabilelor (5) tinde în probabilitate către media comună m a acestor variabile.

Demonstrație. Putem scrie relațiile :

$$M(\bar{x}) = \frac{1}{n}\sum_{k=1}^{n} M(x_k) = \frac{1}{2}(m + m + \ldots + m) = \frac{mn}{n} = m$$

$$D^2(\bar{x}) = \frac{1}{n^2}\sum_{k=1}^{n} D^2(x_k)$$

Deoarece fiecare $D^2(x_k) < L$, rezultă că suma $\sum_{k=1}^{n} D^2(x_k) < nL$, așa încît relația precedentă devine : $D^2(\bar{x}) < \frac{L}{n}$.

Aplicînd variabilei $\bar{x}$, inegalitatea (4'') a lui Cebîșev obținem :

$$P(|\bar{x} - m| < \varepsilon) \geqslant 1 - \frac{D^2(x)}{\varepsilon^2}.$$

Scanned with OKEN Scanner

De unde, înlocuind $D^2(\bar{x})$ prin numărul mai mare $\frac{L}{n}$:

$$P(|\bar{x} - m| < \varepsilon) > 1 - \frac{L}{n\varepsilon^2}.$$

Aşadar, avem inegalităţile :

$$1 - \frac{L}{n\varepsilon^2} < P\left(\left|\frac{x_1 + x_2 + \ldots + x_n}{n} - M\left(\frac{x_1 + x_2 + \ldots + x_n}{n}\right)\right| < \varepsilon\right) \leqslant 1.$$

Pentru $n \to \infty$ obţinem :

$$P\left(\left|\frac{x_1 + x_2 + \ldots + x_n}{n} - M\left(\frac{x_1 + x_2 + \ldots + x_n}{n}\right)\right| < \varepsilon\right) \to 1 \qquad (5_3)$$

adică media $\frac{1}{n}(x_1 + x_2 + \ldots + x_n)$ converge în probabilitate către $M\left(\frac{1}{n}(x_1 + x_2 + \ldots + x_n)\right) = m$.

Frecvenţa relativă. Teorema lui Bernoulli. Să considerăm n probe independente şi evenimentul A care are în fiecare probă aceeaşi probabilitate constantă p de a se realiza.

Dacă x_1 este rezultatul primei probe, evident x_1 este o variabilă aleatorie şi ia valoarea 1 cu probabilitatea p, cînd se produce evenimentul A şi valoarea 0 cu probabilitatea $1 - p = q$ cînd nu se realizează A. În mod analog definim şi variabilele $x_2, x_3, \ldots, x_n$. Numărul de cîte ori se realizează A în cele n probe este desigur egal cu suma $x_1 + x_2 \ldots + x_n \leqslant n$. Prin definiţie, se numeşte *frecvenţă relativă* a evenimentului A în cele n probe, raportul

$$f_n(A) = \frac{x_1 + x_2 + \ldots + x_n}{n} \qquad (6)$$

Valoarea medie a variabilei x_k este :

$$M(x_k) = 1 \cdot p + 0 \cdot q = p \qquad (6_1)$$

iar dispersia

$$D^2(x_k) = M(x_k - p)^2 = (1 - p)^2 p + (0 - p)^2 q = \\ = q^2 p + p^2 q = pq(q + p) = pq \qquad (6_2)$$

Deoarece $q < 1$ rezultă $D^2(x_k) < p$ pentru orice $k = 1, 2, \ldots n$; dispersiile $D^2(x_k)$ sînt deci egal mărginite. În consecinţă varia-

bilele $x_1, x_2, \ldots, x_n$ satisfac condițiile 1° — 3° ale teoremei lui Cebîșev. Relația (5_3) devine în acest caz :

$$P\left(|f_n - M(f_n)| < \varepsilon\right) \underset{n \to \infty}{\longrightarrow} 1$$

Dar :

$$M(f_n) = M\left(\frac{x_1 + x_2 + \ldots + x_n}{n}\right) = \frac{M(x_1) + M(x_2) + \ldots + M(x_n)}{n} = \frac{np}{n} = p.$$

De unde rezultă :

$$P\left(|f_n - p| < \varepsilon\right) \underset{n \to \infty}{\longrightarrow} 1 \qquad (6_3)$$

Am obținut astfel *teorema lui Bernoulli* care afirmă că *frecvența relativă a unui eveniment cu probabilitatea constantă p, în fiecare probă, converge în probabilitate către numărul p.*

Dat fiind că diferența dintre frecvența relativă a unui eveniment, pe care o putem obține în baza probelor făcute, și probabilitatea evenimentului care este un număr necunoscut, poate fi făcută ori cît de mică, cu o probabilitate apropiată de 1 dacă numărul probelor crește, această teoremă poate fi considerată ca una din teoremele fundamentale ale statisticii matematice.

D e f i n i ț i e. Variabila $x(n)$, depinzînd de n, se spune că estimează *absolut corect* numărul a, dacă

$$1°\ M[x(n)] = a$$
$$2°\ M[x(n) - a]^2 \underset{n \to \infty}{\longrightarrow} 0 \qquad (6_4)$$

Aceste două condiții sînt îndeplinite în particular pentru $a = p$, $f_n = x(n)$

Într-adevăr, am demonstrat mai sus că

$$M(f_n) = p \qquad (6_5)$$

Să calculăm acum $M(f_n - p)^2$. Vom avea :

$$M(f_n - p)^2 = M\left(\frac{x_1 + x_2 + \ldots + x_n}{n} - p\right)^2 = M\left(\frac{x_1 + x_2 + \ldots + x_n - np}{n}\right)^2$$

Putem însă scrie relația

$$\left(\frac{x_1 + x_2 + \ldots + x_n - np}{n}\right)^2 = \frac{1}{n^2}\left[(x_1 - p) + (x_2 - p) + \ldots + (x_n - p)\right]^2 =$$

$$= \frac{1}{n^2}\sum_{k=1}^{n}(x_k - p)^2 + \frac{2}{n^2}\sum_{k,h}(x_k - p)(x_h - p).$$

De unde deducem :

$$M(f_n - p)^2 = \frac{1}{n^2} \sum_{k=1}^{n} M(x_k - p)^2 + \frac{2}{n^2} \sum_{k,h} M(x_k - p)(x_h - p)$$

$$D^2(x_k) = M(x_k - p)^2 = pq \text{ după relația} \qquad (6_2)$$

Avem însă : $D^2(x_k - p)(x_h - p) = M(x_k - p) \cdot M(x_k - p) = 0$ întrucît variabilele x_k și x_h sînt independente, iar $M(x_k - p) = M(x_k) - p = 0$ după relația (6_1).

În consecință,

$$M(f_n - p)^2 = D^2(f_n) = \frac{pq}{n} \qquad (6_5)$$

Rezultă de aici că :

$$M(f_n - p)^2 \underset{n \to \infty}{\longrightarrow} 0. \qquad (6_6)$$

așa încît în baza definiției de mai sus putem afirma că *probabilitatea p este estimată absolut corect de frecvența relativă* f_n. Pentru aceste motive în aplicațiile calculului probabilităților se folosește relația aproximativă

$$p \approx f_n \qquad (6_7)$$

Așa, de exemplu, dacă într-un lot de 1 500 de piese, 6 sînt rebuturi, noi spunem că probabilitatea ca o piesă să fie rebut, în condițiile lotului respectiv evident, este egală în medie cu $\frac{6}{1\,500}$.

Teorema asimptotică a lui Poisson este o generalizare a teoremei lui Bernoulli în sensul că evenimentul A are în prima probă probabilitatea p_1, în a doua probabilitatea p_2, ..., în a n-a probă probabilitatea p_n. Variabila x_k reprezentînd rezultatul probei de rang k poate lua valoarea 1, cu probabilitatea p_k dacă se realizează A și valoarea 0 cu probabilitatea $1 - p_k = q_k$ dacă nu se realizează A.

În consecință pentru orice $k = 1, 2, \ldots, n$ avem :

$$M(x_k) = 1 \cdot p_k + 0 \cdot q_k = p_k \qquad (6_8)$$

$$D^2(x_k) = (1 - p_k)^2 + (0 - p_k)^2 q_k = p_k q_k \qquad (6_9)$$

Scanned with OKEN Scanner

După inegalitatea (4'') a lui Cebîșev putem scrie :

$$P\left(\left|\frac{x_1 + x_2 + \ldots + x_n}{n} - M\left(\frac{x_1 + \ldots + x_n}{n}\right)\right| < \varepsilon\right) \geqslant 1 - \frac{D^2\left(\frac{x_1 + \ldots + x_n}{n}\right)}{\varepsilon^2}.$$

Să calculăm acum :

$$M\left(\frac{\sum_{k=1}^{n} x_k}{n}\right) \text{ și } D^2\left(\frac{\sum_{k=1}^{n} x_k}{n}\right)$$

Vom avea :

$$M\left(\frac{\sum_{k=1}^{n} x_k}{n}\right) = \frac{1}{n}\sum_{k=1}^{n} M(x_k) = \frac{1}{n}\sum_{k=1}^{n} p_k \qquad (6'_8)$$

după relația (6_8).

$$D^2\left(\frac{\sum_{k=1}^{n} x_k}{n}\right) = \frac{1}{n^2}\sum_{k=1}^{n} D^2(x_k) = \frac{\sum_{k=1}^{n} p_k q_k}{n} \qquad (6'_9)$$

în baza relațiilor (6_9).

Așadar, putem scrie inegalitatea :

$$P\left(\left|\frac{x_1 + x_2 + \ldots + x_n}{n} - \frac{1}{n}\sum_{k=1}^{n} p_k\right| < \varepsilon\right) \geqslant 1 - \frac{\sum_{k=1}^{n} p_k q_k}{n\,\varepsilon^2} \qquad (6_{10})$$

Introducînd probabilitatea medie

$$p = \frac{1}{n}(p_1 + p_2 + \ldots + p_n) \qquad (6_{11})$$

și ținînd seama de formula (6) putem scrie relația :

$$P(|f_n - p| < \varepsilon) \underset{n \to \infty}{\longrightarrow} 1 \qquad (6'_3)$$

adică aceeași relație pe care am găsit-o în cazul teoremei lui

Scanned with OKEN Scanner

Bernoulli, cu deosebirea că în locul constantei p din relația (6_3) se ia acum media (6_{11}).

Observație. Notînd ca de obicei abaterile

$$x - M(x) = \xi,\ x_k - M(x_k) = \xi_k,\ 1 \leqslant k \leqslant n \qquad (6_{12})$$

putem scrie relația :

$$x - M(x) = x_1 + \ldots + x_n - M(x_1 + \ldots + x_n) =$$
$$= x_1 + \ldots + x_n - M(x_1) - \cdot\cdot - M(x_n)$$

De unde deducem :

$$\xi = \xi_1 + \xi_2 + \ldots + \xi_n \qquad (6_{13})$$

În condițiile de mai sus se demonstrează (acest lucru îl vom face ceva mai tîrziu) că variabila ξ admite pentru $n \to \infty$ o distribuție limită și această distribuție, este distribuția normală.

Teorema lui Poisson a fost extinsă de Cebîșev și școala sa în modul următor :

Fie $p_1, p_2, \ldots, p_n, \ldots$ probabilitățile evenimentului A într-o serie nelimitată de probe independente. Dacă seria $\sum_{k=1}^{\infty} p_k q_k$ este divergentă, abaterea

$$\xi = \xi_1 + \xi_2 + \ldots + \xi_n + \ldots \qquad (6'_{13})$$

admite o distribuție normală.

Teoremele lui Cebîșev, Bernoulli și Poisson, numite *„legile numerelor mari"* exprimă un fapt foarte important din punct de vedere statistic. Dacă dintr-o colectivitate se face o selecție al cărei volum este suficient de mare, mediile empirice vor aproxima caracteristicile teoretice respective — probabilitatea, valoarea medie etc. Cunoașterea acestor caracteristici ajută la descoperirea legilor fundamentale ale procesului cercetat și acesta este tocmai obiectivul principal al oricărui studiu științific al fenomenelor de masă. De obicei prin legea numerelor mari se înțelege ceva mai mult decît teoremele pînă aci enumerate și anume un principiu general în baza căruia acțiunea unui număr mare de factori aleatori conduce, în condiții destul de largi, pe care de altfel le vom preciza ulterior, la un rezultat aproape independent de întîmplare.

Teorema limită centrală Teorema lui Poisson este la rîndul său un caz particular al unei teoreme celebre descoperită de Laplace. În locul abaterilor $\xi_1, \xi_2, \ldots \xi_n$, din

(6_{13}), Laplace își propune să determine distribuția variabilei aleatorii x definită prin relația

$$x = x_1 + x_2 + \ldots + x_n \qquad (6_{14})$$

În concepția lui Laplace, fiecare din variabilele x_k pot lua un număr finit sau o infinitate de valori cu probabilități necunocute. În asemenea condiții, problema aflării distribuției variabilei x, adică găsirea densității sale f(x) sau a funcției de repartiție F(x), pare nedeterminată ; Laplace a admis ipoteza că efectul aleator al neregularităților provenind din variațiile fiecărui termen se neutralizează, așa încît pentru variabila sumă se va obține o distribuție limită, independentă de aceste variații interioare și el a demonstrat că această distribuție limită este cea normală.

Problema pusă de Laplace a fost reluată de Cebîșev care a tratat-o pe o cale cu totul diferită de Laplace și a stabilit totodată și condițiile suficiente pentru valabilitatea rezultatului.

Teorema lui Liapunov. Puțin mai tîrziu, Liapunov, stabilește o condiție suficientă pentru valabilitatea teoremei mai puțin restrictivă decît condițiile lui Cebîșev. Teorema stabilită de el o putem enunța sub o formă ceva mai simplă astfel :

Se consideră șirul infinit de variabile aleatorii x_1, x_2, ..., x_n, ... independente între ele și fie X_n suma parțială de ordinul n :

$$X_n = x_1 + x_2 + \ldots + x_n \qquad (6_{15})$$

Fie η_n abaterea normată a variabilei X_n

$$\eta_n = \frac{X_n - M(X_n)}{D(X_n)} \qquad (6_{16})$$

și $F_n(x)$ funcția sa de repartiție, adică :

$$F_n(x) = P(X_n < x) \qquad (6_{17})$$

Dacă este satisfăcută condiția

$$\lim_{n \to \infty} \frac{M|\xi_1|^3 + M|\xi_2|^3 + \ldots + M|\xi_n|^3}{[D(X_n)]^{3/2}} \to 0 \qquad (6_{18})$$

atunci pentru orice x real avem:

$$\lim_{n \to \infty} F_n(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt \qquad (6_{19})$$

Așadar, cînd condiția (6_{18}) este îndeplinită, suma unui mare număr de variabile independente are aproximativ o distribuție normală. Condiția (6_{18}) impune ca rolul diferiților termeni ai sumei să fie foarte mic atunci cînd numărul variabilelor crește nelimitat.

§ 2. Teoreme de convergență

Teoremele pe care le vom analiza în continuare ne ajută să demonstrăm în mod simplu teoremele descrise mai înainte și la determinarea distribuțiilor limită ale variabilelor aleatorii.

D e f i n i ț i e. Fie șirul de variabile aleatorii independente

$$x_1, x_2 \ldots x_n, \ldots \qquad (7)$$

cu funcțiile generatoare de momente respective

$$m_1(t), m_2(t), \ldots m_n(t), \ldots \qquad (7_1)$$

Dacă pentru orice t finit avem:

$$\lim_{n \to \infty} m_n(t) = m(t) \qquad (7_2)$$

unde m(t) este funcția generatoare a variabilei x, se spune că *șirul (7) converge în repartiție către variabila x.*

Am văzut că funcțiile m(t) și $m_n(t)$ se pot scrie sub forma unei serii întregi de t:

$$m(t) = 1 + m_1 \frac{t}{1!} + m_2 \frac{t^3}{2!} + \ldots + m_k \frac{t^k}{k!} + \ldots \qquad (7_3)$$

$$m_n(t) = 1 + m_{n_1} \frac{t}{1!} + m_{n_2} \frac{t^2}{2!} + \ldots + m_{nk} \frac{t^k}{k!} + \ldots \qquad (7'_3)$$

unde $k = 1, 2, \ldots$ iar m_k și m_{nk} sînt respectiv momentele de ordinul k ale variabilelor x și x_n.

Raza de convergență a seriei (7_3) de pildă, este:

$$R_n = \lim_{k \to \infty} \left| \frac{a_{nk}}{a_{n,k+1}} \right| = \lim_{k \to \infty} k \left| \frac{m_{nk}}{m_{n,k+1}} \right|$$

Scanned with OKEN Scanner

Dacă momentele m_{nk} sînt finite, pentru orice k, atunci $R_n = \infty$, deci seriile (7_3) sau $(7'_3)$ sînt convergente pentru orice t real. Dacă t_0 este fix, pentru orice $|t| < t_0$ aceste serii sînt absolut şi uniform convergente şi pot fi dezvoltate în serie Taylor, adică sînt analitice în orice punct din acest interval.

T e o r e m a I. Dacă variabilele (7) au momente finite de orice ordin şi dacă ele tind în repartiţie către variabila x, atunci momentele $m_{nk} \to m_k$ pentru $n \to \infty$.

Demonstraţie. Funcţiile $m_n(t)$ şi $m(t)$ fiind analitice în intervalul $|t| < t_0$, diferenţa lor este analitică în acelaşi interval, deci, după formula lui Taylor avem:

$$m_n(t) - m(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \left[m_n(t) - m(t) \right]^{(k)}_{t=0}$$

În condiţiile teoremei $m_n(t) - m(t) \to 0$ pentru $n \to \infty$, deci pentru orice $k = 1, 2, \ldots$ va trebui să avem:

$$\left[m_n(t) - m(t) \right]^{(k)}_{t=0} = m_n^k(0) - m^{(k)}_{(o)} \to 0, \text{ cînd } n \to \infty.$$

Întrucît,

$$m_n^{(k)}(0) = m_{nk}$$

$$m^{(k)}(0) = m_k$$

rezultă $m_{nk} \to m_k$ pentru $n \to \infty$ şi $k = 1, 2, \ldots$

T e o r e m a II. Dacă momentele $m_{nk} \to m_k$ atunci $m_n(t) \to m(t)$ pentru $n \to \infty$.

Demonstraţie. Vom putea scrie relaţia:

$$\left| m_n(t) - m(t) \right| < \sum_{k=0}^{\infty} \left| m_{nk} - m_k \right| \frac{|t|^k}{k!}$$

Prin ipoteză, $m_{nk} \to m_k$, deci unui $\varepsilon > 0$ îi corespunde un $k(\varepsilon)$ aşa fel ca pentru $k > k(\varepsilon)$ să avem: $|m_{nk} - m_k| < \varepsilon$. Atunci obţinem în intervalul $|t| < t_0$ $|m_{nk} - m_k| < \varepsilon e^{|t|} < \varepsilon e^{t_0}$ Dacă alegem $\eta > 0$, aşa ca $\varepsilon = \eta^{e-t_0}$ rezultă: $|m_n(t) - m(t)| < \eta$ dacă avem $k > k(\varepsilon)$, pentru orice t din intervalul $(-t_0, +t_0)$.

T e o r e m a III. Dacă funcţiile de repartiţie $F_1(x) \ldots F_n(x) \ldots$ ale variabilelor $x_1, x_2, \ldots x_n, \ldots$ tind către funcţia de repartiţie $F(x)$, funcţiile generatoare $m_1(t), m_2(t), \ldots, m_n(t), \ldots$ converg către funcţia generatoare $m(t)$, corespunzătoare funcţiei $F(x)$.

Scanned with OKEN Scanner

Demonstrație. Deoarece $F_n(x) \to F(x)$ în punctele $x = -a$ și $x = b$ în care $F(x)$ este continuă, alese în așa fel ca : $F(-a) < \varepsilon$, $F(+\infty) - F(b) < \varepsilon$ putem determina un $N(\varepsilon)$ astfel ca :

$$| F_n(-a) - F(-a) | < \varepsilon$$

$$| F_n(b) - F(b) | < \varepsilon$$

dacă $n > N(\varepsilon)$. Înlocuind densitatea $f(x)\,dx$ prin $d\,F(x)$ putem scrie :

$$m_n(t) - m(t) = \int_{-a}^{b} e^{tx}\, d\, F_n(x) - \int_{-a}^{b} e^{tx}\, dF(x) + w_n(t)$$

unde $w_n(t) \to 0$ pentru $n > N(\varepsilon)$.

Urmează acum să determinăm un $N_1(\varepsilon) > N(\varepsilon)$ în așa fel ca diferența integralelor să fie mai mică decît ε.

În același mod se demonstrează și teorema reciprocă.

T e o r e m a IV. Dacă funcțiile generatoare $m_n(t) \to m(t)$, atunci și funcțiile de repartiție $F_n(x) \to F(x)$ pentru $n \to \infty$.

Observație. Teoremele enunțate pînă acum relativ la funcția generatoare, precum și altele pe care nu le mai reproducem aici, sînt evident valabile și pentru funcțiile caracteristice $\varphi_n(t)$ respectiv $\varphi(t)$.

Aplicații. 1°. *Demonstrația teoremei limitei centrale într-un caz particular.*

Variabilele $x_1, x_2, \ldots, x_n$ iau valorile 1 sau 0 cu probabilitățile p și $1 - p = q$, după cum în cele n probe independente evenimentul A se realizează sau nu.

Ne propunem să determinăm distribuția limită a sumei $X_n = x_1 + x_2 + \ldots + x_n$, pentru n crescînd nelimitat.
În acest scop vom cerceta distribuția variabilei normate

$$\eta_n = \frac{X_n - M(X_n)}{D(X_n)}$$

Deoarece x_k ia valorile 1 sau 0 cu probabilitățile p și q avem :

$$M(x_k) = p, \quad D^2(x_k) = p\,q, \quad m_k(t) = pe^t + q$$

de unde rezultă : $M(X_n) = np$, $D^2(X_n) = npq$ și cum funcția generatoare a sumei X_n este egală cu produsul funcțiilor generatoare ale termenilor, putem scrie :

$$m_{X_n}(t) = \prod_{k=1}^{n} m_k(t) = (p\,e^t + q)^n$$

Scanned with OKEN Scanner

Atunci funcția generatoare a abaterii normate va fi

$$m_{\eta_n}(t) = e^{-\frac{npt}{\sqrt{npq}}}\left(p\,e^{\frac{t}{\sqrt{npq}}} + q\right)^n$$

sau încă :

$$m_{\eta_n}(t) = \left[e^{-\frac{pt}{\sqrt{npq}}}\left(p\,e^{\frac{t}{\sqrt{npq}}} + q\right)\right]^n = \left[p\,e^{\frac{qt}{\sqrt{npq}}} + qe^{-\frac{pt}{\sqrt{npq}}}\right]^n =$$

$$= \left[\left(p + \frac{pqt}{\sqrt{npq}} + \frac{pq^2t^2}{2\,npq} + \frac{pq^3t^3}{6(npq)^{3/2}} + \cdots\right) + \right.$$

$$\left. + \left(q - \frac{pqt}{\sqrt{npq}} + \frac{p^2qt^2}{2\,npq} - \frac{p^3\,qt^3}{6\,(npq)^{3/2}} + \cdots\right)\right]^n =$$

$$= \left[1 + \frac{t^2}{2n} + 0\left(\frac{1}{n^{3/2}}\right)\right]^n$$

unde $0\left(\frac{1}{n^{3/2}}\right)$ reprezintă suma termenilor neglijați în care exponenții lui n sînt mai mari decît 3/2, deci $0\left(\frac{1}{n^{3/2}}\right) \to 0$.

Folosind relația :

$$\lim_{n \to \infty}\left(1 + \frac{\alpha}{n} + \frac{\beta}{n^2} + \cdots\right) = e^{\alpha}$$

se găsește :

$$\lim_{n \to \infty} m_{\eta_n}(t) = \lim_{n \to \infty}\left[1 + \frac{\left(\frac{t^2}{2}\right)}{n} + 0\left(\frac{1}{n^{3/2}}\right)\right]^n = e^{\frac{t^2}{2}}$$

Așadar, funcția generatoare a variabilei η_n tinde către funcția generatoare a legii normale, de unde rezultă în baza teoremei I, că variabila X_n admite o distribuție limită normală.

De aici deducem egalitatea apropiată

$$P(a < \eta_n < b) \approx \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{t^2}{2}}\,dt = \Phi(b) - \Phi(a) \qquad (7_1)$$

unde $\Phi(t)$ este funcția cunoscută a lui Laplace

$$\Phi(t) = \frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{x^2}{2}}\,dx \qquad (7_2)$$

Dacă luăm $a = -k$, $b = +k$ obținem probabilitatea

$$P(-k < \eta_n < +k) = \frac{1}{\sqrt{2\pi}} \int_{-k}^{+k} e^{-\frac{t^2}{2}} dt \qquad (7_6)$$

și deoarece inegalitățile $-k < \eta_n < +k$ sînt echivalente cu inegalitatea $|\eta_n| < k$, iar

$$\Phi(-k) = -\Phi(k) \qquad (7_7)$$

se găsește :

$$P(|\eta_n| < k) = 2\,\Phi(k) \qquad (7_8)$$

iar valorile $\Phi(k)$ se citesc din tabele.

Din relația (7_6) avem de asemenea probabilitatea

$$P(np - k\sqrt{npq} < X_n < np + \sqrt{npq}) = \frac{1}{\sqrt{2\pi}} \int_{-k}^{+k} e^{-\frac{t^2}{2}} dt \qquad (7'_6)$$

În aplicații se folosește de obicei formula (7_4). În acest scop se calculează $M(x)$ și σ și se formează apoi abaterea normată $\eta = [x - M(x)]/\sigma$.

Exemplu. Probabilitatea ca un fuzeu electric să fie în afara zonei de control s-a calculat în baza unor probe nu meroase egală cu 0,2. Dacă se extrage o selecție de 400 bucăți, după schema bilei revenite să se calculeze probabilitatea ca în afara zonei de control să se găsească între 70 și 100 fuzee.

Variabila x reprezentînd numărul fuzeelor în afara zonei de control are o distribuție binomială care se încadrează în teorema limitei centrale de mai sus. Vom avea deci :

$$M(x) = np = 400 \cdot 0{,}2 = 80$$

$$D^2(x) = npq = 400 \cdot 0{,}2 \cdot 0{,}8 = 64 \text{ deci } \sigma = 8$$

Deoarece se cere ca $70 < x < 80$ avem limitele abaterii normate :

$$\frac{70 - np}{\sigma} < \frac{x - np}{\sigma} < \frac{100 - np}{\sigma}$$

sau încă $-1{,}25 < \eta < 2{,}50$

Aplicînd (7_4) și ținînd seama de relația (7_7) găsim : $P(-1{,}25 < \eta < 2{,}50) = P(70 < x < 100) = \Phi(2{,}50) + \Phi(1{,}25)$

În tabele se găsesc valorile : $\Phi(2{,}50) = 0{,}4938$, $\Phi(1{,}25) = 0{,}3944$, deci probabilitatea căutată este egală cu 0,8882.

Scanned with OKEN Scanner

2. *Demonstrația teoremei lui Liapunov* (Teorema limită centrală). Enunțul complet al acestei teoreme este următorul :

Fie $x_1, x_2, \ldots, x_n, \ldots$ un șir de variabile aleatorii independente și sumele parțiale

$$X_n = x_1 + x_2 + \ldots + x_n \quad (n = 1, 2, \ldots)$$

Să presupunem că pentru orice k avem valorile finite

$$M(x_k) = m_k ; \; M(x_k - m_k)^2 = \mu_k, \; M|x_k - m_k|^3 = \rho_k^3$$

Dacă însemnăm :

$$m_{(n)} = \sum_{k=1}^{n} m_k ; \; \mu_{(n)}^2 = \sum_{k=1}^{n} \mu_k^2 ; \; \rho_{(n)}^3 = \sum_{k=1}^{n} \rho_k^3$$

și dacă :

$$\lim_{n \to \infty} \frac{\rho_{(n)}}{\mu_{(n)}} = 0$$

atunci șirul de variabile $\{x_{(n)}\}$ tinde la limită către distribuția normală.

Vom reda demonstrația acestei teoreme cu ajutorul funcției caracteristice. Fie $f_k(t)$ funcția caracteristică a abaterii $(x_k - m_k)$, adică :

$$f_k(t) = M(e^{it(x_k - m_k)}) = 1 - \frac{1}{2}\mu_k^2 t^2 + \frac{1}{6}\Phi \rho_k^3 t^3 \quad 0 < \theta < 1$$

Schimbînd t prin $t/\mu_{(n)}$ obținem :

$$f_k\left(\frac{t}{\mu_{(n)}}\right) = 1 - \frac{1}{2}\frac{\mu_k}{\mu_{(n)^2}} t^2 + \frac{1}{6}\theta \frac{\rho_k^3}{\mu_{(n)^2}} t^3$$

de unde :

$$\ln f_k\left(\frac{t}{\mu_{(n)}}\right) = \log(1 + \alpha)$$

în care am notat :

$$\alpha = -\frac{\sigma_k^2 t^2}{2\sigma^2} + \theta \frac{\rho_k^3 t^3}{6\sigma^3} ; \; \sigma_k^2 = \mu_k : \mu_{(n)}^2 = \sigma^2$$

Avem însă :

$$\frac{\rho_k}{\mu_{(n)}} < \frac{\rho_{(n)}}{\mu_{(n)}} < 1$$

Scanned with OKEN Scanner

$\mu_k \leqslant \rho$. P utem deci scrie pe α sub forma:

$$\alpha = \frac{\rho_k^2}{\mu_{(n)}^2}\left(\frac{t^2}{2} + \frac{|t|^3}{6}\right), \quad |\theta| < 1$$

Dacă $n \to \infty$, $\frac{\rho_k}{\mu_{(n)}} \to 0$, deci putem lua n suficient de mare pentru ca $|\alpha| < 1/2$. Avem dezvoltarea cunoscută:

$$\ln(1+\alpha) = \alpha - \frac{\alpha^2}{2} + \frac{\alpha^3}{3} \ldots = \alpha - \frac{\alpha^2}{2}\left(1 - \frac{2}{3}\alpha + \frac{3}{4}\alpha^2 \ldots\right) =$$

$$= \alpha + \frac{1}{2}\theta\alpha^2\left(1 + \frac{1}{2} + \frac{1}{2^2} + \ldots\right) = \alpha + \theta\alpha^2$$

În consecință, vom putea scrie:

$$\ln f_n\left(\frac{t}{\mu_{(n)}}\right) = -\frac{\sigma_k^2}{\mu_{(n)}^2} \cdot \frac{t^2}{2} + \theta\frac{\rho_k^2}{\mu_{(n)}^2}\left[\frac{1}{6}|t|^3 + \left(\frac{1}{2}t^2 + \frac{1}{6}|t|^3\right)^2\right]$$

Funcția caracteristică a variabilei $x_n - m_{(n)}$, va fi:

$$F_n(t) = f_1(t) \cdot f_2(t) \ldots f_n(t)$$

și deci:

$$\ln F_{(n)}\left(\frac{t}{\mu_{(n)}}\right) = \sum \ln f_k\left(\frac{t}{\mu_{(n)}}\right) = -\frac{t^2}{2} + \theta\frac{\rho_n^3}{\mu_{(n)}^3}\left[\frac{1}{6}|t|^3 + \right.$$

$$\left. + \left(\frac{1}{2}t^2 + \frac{1}{6}|t|^3\right)^2\right]$$

Pentru $n \to \infty$, rezultă limita:

$$\lim_{n\to\infty} \ln F_n\left(\frac{t}{\mu_{(n)}}\right) = -\frac{t^2}{2}$$

Legea normală tip avînd funcția caracteristică egală cu $e^{-\frac{t^2}{2}}$, rezultă că variabila X_n admite ca distribuție limită, distribuția normală[1].

---

[1] Pentru un studiu mai detaliat al teoremelor din ultima parte a acestui capitol pot fi consultate cu folos următoarele tratate din care am reprodus o parte a demonstrațiilor din text: *O. Onicescu*, „Calculul probabilităților", Ed. tehnică. 1956, Buc. *Gh. Mihoc*, „Elemente de calculul probabilităților" Ed. tehnică, 1954, București.

Scanned with OKEN Scanner

Probleme

1. Să se arate că distribuția lui Poisson tinde către distribuția normală cînd parametrul său crește nelimitat.

2. Pentru valori mari ale lui n, distribuția hipergeometrică este puțin diferită de cea normală.

3. Dacă numărul gradelor de libertate crește nelimitat distribuția lui Student se apropie de cea normală.

4. Se dă șirul de variabile independente $x_1, x_2 \ldots x_n$ pentru care

$$M(x_i) = 0,\ M(x_i^2) = 1,\ 1 \leq i \leqslant n.$$

Să se determine distribuția variabilei $U_n = \Sigma x_n = \sqrt{\Sigma x_n^2}$ cînd n crește nelimitat.

Scanned with OKEN Scanner

CAPITOLUL IX

SELECȚII ALEATORII

§ 1. Distribuții de selecție

Să considerăm o populație Γ din care să extragem o selecție de mărime n care ne conduce la valorile

$$x_1\ x_2,\ \ldots,\ x_n \qquad (1)$$

Prin *selecție aleatorie* se înțelege că valoarea x_1 obținută în prima probă nu are nici o influență asupra valorii x_2 obținută în proba a doua, deci variabilele x_1 și x_2 sînt variabile independente așa cum au fost definite în capitolele precedente; același lucru se spune despre variabilele x_3, x_4, ...x_n.

Din definiția dată mai sus, selecțiile aleatorii se referă la populațiile continue (infinite), precum și la populațiile finite în care probele se execută după schema bilei revenite (selecții repetate).

Presupunem că în populația Γ s-a definit o variabilă aleatorie x a cărei densitate de probabilitate este f (x). Din această populație facem o selecție de mărime n, care ne dă valorile (1). Deoarece fiecare din variabilele x_1, x_2,.., x_n pot lua oricare din valorile lui x, *vom considera aceste n variabile ca niște variabile aleatorii identice cu x.*

Mulțimea de valori (x_1, x_2, ..., x_n) reprezintă în spațiul n-dimensional un punct avînd aceste numere drept coordonate. Mulțimea tuturor sistemelor posibile de n asemenea valori formează o populație cu n variabile sau o distribuție n-dimensională. Ne propunem *să determinăm această distribuție în funcție de densitatea f (x) a variabilei x definită în M.*

Densitatea distribuției cu n variabile pe care s-o însemnăm prin $f(x_1, x_2, \ldots, x_n)$ va trebui să satisfacă relația

$$P(a_1<x_1<b_1,\ a_2<x_2<b_2,\ \ldots\ a_n<x_n<b_n) =$$

$$= \int_{a_1}^{b_1}\int_{a_2}^{b_2}\ldots\int_{a_n}^{b_n} f(x_1, x_2, \ldots x_n)\, dx_1 \ldots dx_n \qquad (2)$$

pentru orice valori ale numerelor a_h, b_h $(1 \leqslant h \leqslant n)$.

În baza definiției selecției aleatorii, variabilele $x_1, x_2, \ldots x_n$ sînt independente, deci densitatea distribuției este egală cu produsul densităților marginale ale variabilelor. Deoarece densitatea marginală a lui x_1 este $f(x_1)$ căci x_1 este identic cu x, a lui x_2 este $f(x_2)$, ... putem scrie relația

$$f(x_1, x_2, \ldots x_n) = f(x_1) \cdot f(x_2) \ldots (x_n) \qquad (2_1)$$

sau încă relația echivalentă:

$$P(a_1<x_1<b_1,\ a_2<x_2<b_2) \ldots a_n<x_n<b_n) =$$

$$= P(a_1<x_1<b_1) \cdot P(a_2<x_2<b_2) \ldots P(a_n<x_n<b_n) \qquad (2_2)$$

Exemplu. Variabila x definită în Γ poate lua valorile 1 și 0 cu probabilitățile p și q (distribuția binomială). Pentru fiecare probă avem densitatea:

$$f(x) = C_1^x\, p^x\, q^{1-x},\quad x = 0,\ 1$$

sau încă, ținînd seama că $C_1^0 = C_1^1 = 1$

$$f(x) = p^x \cdot q^{1-x},\qquad x = 0,\ 1$$

Dacă de exemplu populația Γ este formată din mulțimea pieselor produse de o mașină într-o oră și 1 corespunde unei piese standard, iar 0 unei piese rebut cu probabilitățile respective $p = 0{,}997$ și $q = 0{,}003$, o selecție de trei piese ne conduce la densitatea:

$$f(x_1, x_2, x_3) = 0{,}997^{x_1+x_2+x_3} \cdot 0{,}003^{3-x_1-x_2-x_3},\ x_1, x_2, x_3 = 0,1$$

Scanned with OKEN Scanner

Această funcție este definită în cele opt puncte vîrfuri ale cubului de latură 1, așezat ca în figura de mai jos :

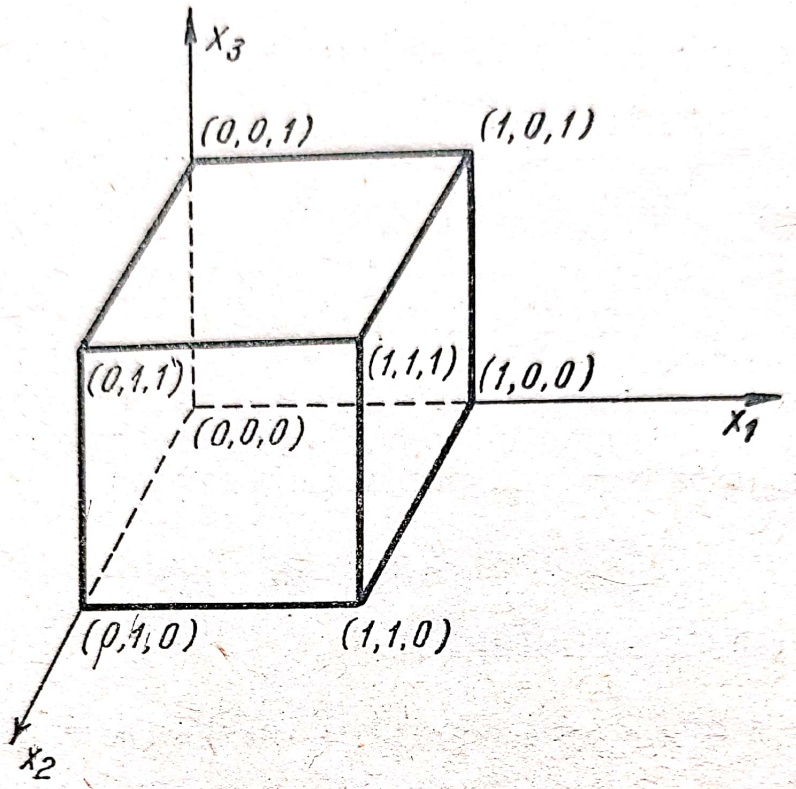

Fig. 48

Probabilitatea ca toate piesele să fie standard se va obține evident făcînd în formula de mai sus $x_1 = x_2 = x_3 = 1$, deci : $f(1, 1, 1) = 0{,}997^3$

Se verifică ușor condiția :

$$\sum_{x_1=0}^{1} \sum_{x_2=0}^{1} \sum_{x_3=0}^{1} f(x_1, x_2, x_3) = 1$$

§ 2. Momentele selecției. Am definit în capitolul I momentul de ordinul k al selecției. Presupunînd că valorile date de selecție sînt diferite între ele, avem :

$$m'_k = \frac{1}{n} \sum_{h=1}^{n} x_h^k, \quad k = 1, 2, \ldots \tag{3}$$

Momentul de primul ordin m_1 se notează, după cum știm, de obicei, prin $\bar{x}$ și poartă numele de medie de selecție, deci

$$\bar{x} = \frac{1}{n} \sum_{h=1}^{n} x_h. \tag{3_1}$$

Am menționat în paginile precedente faptul important pentru teoria selecției și anume că *valorilor* (*I*) *le putem da o dublă interpretare*: [1° După efectuarea selecției ele reprezintă anumite mărimi concrete utile pentru concluziile ce interesează seria de n probe făcută și [2° Înainte de efectuarea probelor ele pot lua toate valorile posibile ale lui x, deci sînt identice cu aceasta, așa încît pentru orice h = 1, 2, ... n și k = 1, 2,... vom avea

$$M(x_h^k) = M(x^k) = m_k$$

$$M(x_h - m)^k = M(x - m)^k = \mu_k. \qquad (3_2)$$

Media de selecție. Deoarece $\bar{x}$ este o sumă de n variabile aleatorii este și ea o variabilă aleatorie și în consecință are o anumită distribuție pe care ne propunem s-o cercetăm în cele ce urmează. Din relația (3_1) deducem valoarea medie a mediei de selecție:

$$M(\bar{x}) = \frac{1}{n} M\left(\sum_{h=1}^{n} x_h\right) = \frac{1}{n} \sum_{h=1}^{n} M(x_h)$$

Dar din (3_2) deducem pentru k = 1 și orice h = 1, 2,...n

$$M(x_h) = M(x) = m_1 = m.$$

De unde rezultă

$$M(\bar{x}) = m \qquad (3_3)$$

Așadar, *valoarea probabilă a mediei de selecție este egală cu valoarea medie a populației* Γ *din care am extras selecția.*

Din punct de vedere practic aceasta înseamnă că dacă facem o serie suficient de mare de s grupe de cîte n probe (adică s selecții) și obținem mediile $\bar{x}_1, \bar{x}_2, \ldots \bar{x}_s$, în baza teoremei lui Cebîșev, media lor $\frac{1}{s}(\bar{x}_1 + \bar{x}_2 + \ldots + \bar{x}_s)$ este aproximativ egală cu media m a populației care, de obicei, este necunoscută și vrem s-o estimăm prin selecție.

Dispersia mediei de selecție va fi:

$$D^2(\bar{x}) = M\left(\frac{1}{n}\sum_{h=1}^{n} x_h - m\right)^2 = M\left(\frac{\sum_{h=1}^{n} x_h - nm}{n}\right)^2 =$$

$$= \frac{1}{n^2} M\left[\sum_{h=1}^{n} (x_h - m)\right]^2 = \frac{1}{n^2} M\left[\sum_{h=1}^{n} (x_h - m)^2 + 2\sum_{h,j} (x_h - m)(x_j - m)\right] =$$

$$= \frac{1}{n^2} \sum_{h=1}^{n} M(x_h - m)^2 + 2\sum_{h,j} M(x_h - m)(x_j - m)$$

Scanned with OKEN Scanner

Din relația (3_2) însă avem :

$$M\ (x_h - m)^2 = M(x - m)^2 = D^2\,(x) = \sigma^2$$

Pe de altă parte, variabilele x_h și x_j fiind independente, putem scrie : $M\ (x_h - m)\ (x_j - m) = M\ (x_h - m).\ M\ (x_j - m) = 0$ deoarece $M\ (x_h - m) = M\ (x_h) - m = m - m = 0$

Notînd dispersia mediei de selecție cu

$$D^2\,(\bar{x}) = \sigma^2_{\bar{x}} \qquad (3_4)$$

obținem relația :

$$\sigma^2_{\bar{x}} = \frac{1}{n^2}\,(\sigma^2 + \sigma^2 + \ldots + \sigma^2)$$

sau încă :

$$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}. \qquad (3_5)$$

Prin urmare, *media de selecție are o împrăștiere de n ori mai mică decît variabila x.*

În baza rezultatelor (3_3) și (3_5) putem scrie :

$$M\ (\bar{x}) = m,\ M\ (\bar{x} - m)^2 \to 0 \qquad (3_6)$$
$$n \to \infty.$$

Așadar, *media de selecție* $\bar{x}$ *estimează absolut corect media m a populației* Γ, așa încît, din punct de vedere practic, cu cît selecția este mai mare cu atît sîntem mai siguri că valoarea mediei de selecție este o bună aproximație a valorii medii a populației din care am extras selecția.

La acest rezultat se putea ajunge și altfel, folosind inegalitatea lui Cebîșev. Avem $M\ (\bar{x}) = m$, $D^2\ (\bar{x}) = \frac{\sigma^2}{n}$, de unde rezultă

$$P\ (|\bar{x} - m| < \varepsilon) > 1 - \frac{D^2\,(\bar{x})}{\varepsilon^2} = 1 - \frac{\sigma^2}{n\,\varepsilon^2}$$

și pentru n suficient de mare

$$\lim_{n \to \infty} P\ (|\bar{x} - m| < \varepsilon) \to 1.$$

Prin urmare, dacă alegem ε un număr destul de mic și determinăm un interval mic în jurul mediei m a populației Γ, vom

putea găsi pe n suficient de mare pentru a obține o valoare apropiată oricît dorim, de 1, a probabilității ca valoarea medie a selecției să se găsească în intervalul ales.

Exemplu. Populația Γ are o valoare medie m necunoscută și dispersia 1. Să se determine mărimea n a selecției așa fel încît, cu o probabilitate de cel puțin 0,97 valoarea medie a selecției să nu difere cu mai mult de 0,5 de valoarea medie a populației.

Cu datele problemei avem : $\sigma^2 = 1$, $\varepsilon = 0{,}5$. Va trebui să calculăm n așa fel ca :

$$1 - \frac{\sigma^2}{n\varepsilon^2} = 1 - \frac{1}{n \times 0{,}5^2} = 0{,}97$$

De unde rezultă $n \approx 134$ probe.

Teoremă. Dacă populația Γ are dispersia σ^2 finită și valoarea medie m, distribuția mediei de selecție se apropie de distribuția normală cu dispersia $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$ și media $M(\bar{x}) = m$, o dată cu creșterea mărimii n a selecției.

Cu alte cuvinte, dacă $g(\bar{x})$ este densitatea de probabilitate a mediei de selecție $\bar{x}$, avem pentru n mare :

$$g(\bar{x}) = \frac{\sqrt{n}}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{n}{2}\left(\frac{\bar{x}-m}{\sigma}\right)^2} \qquad (4)$$

Demonstrația I. Va fi suficient să arătăm că funcția generatoare a variabilei normate

$$\bar{\eta} = \frac{\bar{x} - M(\bar{x})}{D(\bar{x})} = \frac{\bar{x} - m}{\left(\frac{\sigma}{\sqrt{n}}\right)}$$

tinde către funcția generatoare $m_2(t) = e^{\frac{1}{2}t^2}$ a distribuției normale tip.

Funcția generatoare a variabilei $\bar{\eta}$ este :

$$\bar{m}(t) = M\left(e^{t\frac{\bar{x}-m}{\sigma/\sqrt{n}}}\right) = \int_{-\infty}^{+\infty} e^{t\frac{\bar{x}-m}{\sigma/\sqrt{n}}}\, g(\bar{x})\, d\bar{x}$$

Scanned with OKEN Scanner

Avem însă relațiile :

$$\bar{x} - m = \frac{1}{n} \sum_{h=1}^{n} x_h - m = \frac{1}{n} \sum_{h=1}^{n} (x_h - m),$$

deci :

$$M\left(e^{t \frac{\bar{x}-m}{\sigma/\sqrt{n}}} \right) = M\left(e^{t \frac{\frac{1}{n}\sum (x_h - m)}{\sigma/\sqrt{n}}} \right) = M\left(e^{\frac{t}{\sqrt{n}} \sum_h \frac{x_h - m}{\sigma}} \right)$$

Prin urmare vom putea scrie, după definiția funcției generatoare :

$$\bar{m}(t) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} e^{\frac{t}{\sqrt{n}} \Sigma \frac{x_h - m}{\sigma}} f(x_1, \dots x_n)\, dx_n \dots dx_1 =$$

$$= \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} e^{\frac{t}{\sqrt{n}} \sum_{h=1}^{n} \frac{x_h - m}{\sigma}} \prod_{h=1}^{n} f(x_n) \prod_{h=1}^{n} dx_h$$

Integrala multiplă se poate scrie ca un produs de n integrale simple, așa încît,

$$\bar{m}(t) = \int_{-\infty}^{+\infty} e^{\frac{t}{\sqrt{n}} \frac{x_1 - m}{\sigma}} f(x_1)\, dx_1 \int_{-\infty}^{+\infty} e^{\frac{t}{\sqrt{n}} \frac{x_2 - m}{\sigma}} f(x_2)\, dx_2 \dots$$

$$\dots \int_{-\infty}^{+\infty} e^{\frac{t}{\sqrt{n}} \cdot \frac{x_n - m}{\sigma}} f(x_n)\, dx_n). \qquad (4_1)$$

Funcția generatoare a variabilei $\frac{x - m}{\sigma}$ fiind :

$$m_2(t) = \int_{-\infty}^{+\infty} e^{t \frac{x - m}{\sigma}} f(x)\, dx \qquad (4_2)$$

și cum fiecare din variabilele x_h sînt identice cu x avem :

$$\int_{-\infty}^{+\infty} e^{\frac{t}{\sqrt{n}} \cdot \frac{x_h - m}{\sigma}} f(x_h)\, dx_h = m_2\left(\frac{t}{\sqrt{n}}\right) \qquad (4_3)$$

Scanned with OKEN Scanner

Înlocuind în relația (4_1) obținem :

$$\bar{m}(t) = \left[m_2\left(\frac{t}{\sqrt{n}}\right)\right]^n = \left[\left[M\left(e^{\frac{t}{\sqrt{n}} \cdot \frac{x-m}{\sigma}} \right)\right]^n =\right.$$

$$= \left[M\left(1 + \frac{t}{\sqrt{n}} \cdot \frac{x-m}{\sigma} + \frac{t^2}{2n} \cdot \frac{(x-m)^2}{\sigma^2} + \frac{t^3}{6\, n^{3/2}} \cdot \frac{(x-m)^2}{\sigma^3} + \ldots \right)\right]^n .$$

Ținînd seama că $M(x-m)^k = \mu_k$ se găsește :

$$\bar{m}(t) = \left[1 + \frac{\mu_1}{\sigma} \cdot \frac{t}{\sqrt{n}} + \frac{\mu_2}{2\sigma^2} \cdot \frac{t^2}{n} + \frac{\mu_3}{6\sigma^3} \cdot \frac{t^2}{n^{3/2}} + \ldots \right]^n$$

Dar $\mu_1 = 0$, $\mu_2 = \sigma^2$, deci :

$$\bar{m}(t) = \left[1 + \frac{\left(\frac{t^2}{2}\right)}{n} + 0\left(\frac{1}{n^{3/2}}\right)\right]^n$$

unde funcția $0\left(\frac{1}{n^{3/2}}\right)$ conține la numitor pe n cu puterea cea mai mică egală cu $\frac{3}{2}$. Este evident că $0\left(\frac{1}{n^{3/2}}\right) \to 0$, pentru $n \to \infty$ deci :

$$\lim_{n\to\infty} \bar{m}(t) = \lim_{n\to\infty} \left[1 + \frac{\left(\frac{t^2}{2}\right)}{n}\right]^n = e^{\frac{1}{2}t^2} \qquad (4_4)$$

Prin urmare, *media de selecție* $\bar{x}$ *are o distribuție normală cînd volumul selecției crește nelimitat*, oricare ar fi densitatea f(x) a variabilei x definită pe populația Γ.

Observație. Demonstrația dată mai sus presupune existența funcției generatoare de momente a variabilei x. Ea este aceeași și pentru variabilele de tip discret în care caz integralele se înlocuiesc prin sume.

Demonstrația II. Această a doua demonstrație se bazează pe teorema lui Liapunov. Am văzut că dacă avem $x_{(n)} = x_1 + \ldots + x_n$ în care variabilele $x_1, x_2, \ldots, x_n$ sînt independente și dacă

$$\lim_{n\to\infty} \frac{\sum_{h=1}^{n} M\left| x_h - M\, x_h \right|^3}{\sigma^{3/2}_{x_{(n)}}} \to 0 \qquad (4_5)$$

unde $\sigma^2_{x(n)}$ este dispersia sumei $x_{(n)}$, atunci variabila $x_{(n)}$ admite, pentru n suficient de mare o distribuție normală.

Să presupunem că variabila x definită pe Γ are o dispersie σ^2 finită. Variabilele $x_1, x_2, \ldots x_n$, sînt independente și au aceeași distribuție. Am văzut că $M(x_h) = m = M(x)$, $D^2(x_h) = \sigma^2 = D^2(x)$ pentru orice $h = 1, 2, \ldots, n$.

Media de selecție este egală cu suma a n variabile aleatorii independente :

$$\bar{x} = \frac{x_1}{n} + \frac{x_2}{n} + \ldots + \frac{x_n}{n}.$$

Avem :

$$M\left(\frac{x_h}{n}\right) = \frac{1}{n} M(x_h) = \frac{m}{n}; \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}. \tag{4_5}$$

Abaterile :

$$\frac{x_h}{n} - M\left(\frac{x_h}{n}\right) = \frac{x_h}{n} - \frac{m}{n}$$

sînt finite și fie A cel mai mic număr pozitiv pentru care $\left|\frac{x_h}{n} - \frac{m}{n}\right| < A$, oricare ar fi h.

Vom avea :

$$M\left|\frac{x_h}{n} - \frac{m}{n}\right|^3 \leqslant M\left|\frac{x_h}{n} - \frac{m}{n}\right|\left(\frac{x_h}{n} - \frac{m}{n}\right)^2 \leq A\, M\left(\frac{x_h}{n} - \frac{m}{n}\right)^2$$

sau încă :

$$M\left|\frac{x_h}{n} - \frac{m}{n}\right|^3 \leq A\, D^2\left(\frac{x_h}{n}\right) = \frac{A}{n^2} D^2(x_h) = \frac{A\sigma^2}{n^2}.$$

De unde :

$$\Sigma\left|\frac{x_h}{n} - \frac{m}{n}\right|^3 \leq \frac{n\, A\sigma^2}{n^2} = \frac{A\sigma^2}{n}.$$

Această sumă este numărătorul expresiei (4_5)

Numitorul va fi egal cu : $\sigma_{\bar{x}}^{3/2} = \left(\frac{\sigma}{\sqrt{n}}\right)^{3/2} = \frac{\sigma^{3/2}}{n^{3/4}}$, deci :

$$\lim_{n \to \infty} \frac{\frac{A\sigma^2}{n}}{\frac{\sigma^{3/2}}{n^{3/4}}} = \lim_{n \to \infty} \frac{A\sqrt{\sigma}}{\sqrt[4]{n}} = 0$$

Scanned with OKEN Scanner

Condiția (4_5) fiind îndeplinită, suma (4_6), adică media de selecție admite o distribuție normală, oricare ar fi densitatea f(x).

Distribuția momentelor de selecție. Momentul de ordinul k al variabilei x este:

$$m_k = \int_{-\infty}^{+\infty} x^k f(x)\, dx$$

O expresie empirică a acestuia, obținută prin selecție este:

$$m'_k = \frac{1}{n} \sum_{h=1}^{n} x_h^k$$

Acest moment este o variabilă aleatorie, cu valoarea medie

$$M(m'_k) = \frac{1}{n} \sum_{h=1}^{n} M(x_h^k) = \frac{1}{n} \sum_{h=1}^{n} m_k = m_k. \qquad (4_7)$$

Dispersia acestui moment va fi:

$$D^2(m'_k) = M(m'_k - m_k)^2 = M\left(\frac{\sum_{h=1}^{n} x_h^k}{n} - m_k\right)^2 =$$

$$= M\left[\frac{\Sigma (x_h^k - m_k)}{n}\right]^2 = \frac{1}{n^2} M\left[\sum_h (x_h^k - m_k)^2 + 2 \sum_{h,j} (x_h^k - m_k)(x_j^k - m_k)\right] =$$

$$= \frac{1}{n^2} \sum_h M(x_h^k - m_k)^2$$

deoarece $M(x_h^k - m_k) = M x_h^k - m_k = m_k - m_k = 0$

Pe de altă parte avem:

$$M(x_h^k - m_k)^2 = M(x_h^{2k} - 2 m_h^k m_k + m_k^2) = m_{2k} - m_k^2$$

de unde:

$$D^2(m'_k) = \frac{m_{2k} - m_k^2}{n}. \qquad (4_8)$$

În baza relațiilor (4_7) și (4_8) deducem că *momentul de selecție* m'_k *estimează absolut corect momentul teoretic* m_k.

Deoarece m'_k este suma variabilelor independente x_h^k ($1 \leqslant h \leqslant n$), deducem, în baza teoremei lui Liapunov că m'_k

Scanned with OKEN Scanner

admite, pentru n suficient de mare, o distribuție normală cu media m_k *și dispersia* $\frac{m_{2k} - m_k^2}{n}$, adică :

$$f(m'_k) = \frac{1}{\sqrt{2\pi}\,\sigma_{m'k}} \cdot e^{-\frac{1}{2}\left[\frac{m'_k - M(m'_k)}{\sigma_{m'_k}}\right]^2} dm'_k \qquad (4_9)$$

unde :

$$M(m'_k) = m_k, \quad \sigma_{m'k} = \sqrt{\frac{m_{2k} - m_k^2}{n}}. \qquad (4_{10})$$

Determinarea distribuției momentelor centrate de selecție $\mu'_k = \frac{1}{n}\sum_{h=1}^{n}(x_h - \bar{x})^k$ este ceva mai complicată, deoarece valorile x_h se raportează la media de selecție $\bar{x}$ care ea însăși este o variabilă aleatorie. Vom calcula numai valoarea medie a primelor momente empirice centrate folosite des în teoria selecției.

Din relația (6_{11}), capitolul I, deducem :

$$\begin{aligned} \mu'_2 &= m'_2 - m'^2_1 \\ \mu'_3 &= m'_3 - 3\,m'_1\,m'_2 + 2\,m'^3_1 \\ \mu'_4 &= m'_4 - 4\,m'_1\,m'_3 + 6\,m^2_1\,m^1_2 - 3\,m'^4_1. \end{aligned} \qquad (5)$$

Din prima relație se găsește valoarea probabilă :

$$M(\mu'_2) = M(m'_2) - M(m'^2_1).$$

Avem însă :

$$M(m'_2) = m_2 \quad \text{(relația } 4_7 \text{ pentru } k = 2)$$

$$M(m'^2_1) = M(\bar{x}^2) = M\left(\frac{\Sigma x_h}{n}\right)^2 = \frac{1}{n^2} M\left(\sum_h x_h^2 + 2\,\Sigma\, x_h\, x_j\right) =$$

$$= \frac{1}{n^2}\left[\sum_h M(x_h^2) + 2\Sigma\, M(x_h)\cdot M(x_j)\right] = \frac{1}{n^2}\left[n \cdot m_2 + 2\ C_n^2 m_1^2\right] =$$

$$= \frac{m_2}{n} + \frac{n-1}{n} m_1^2.$$

Scanned with OKEN Scanner

Prin urmare,

$$M(\mu_2') = \frac{n-1}{n}(m_2 - m_1^2)$$

sau încă ținînd seama că $m_2 - m_1^2 = \mu_2 = \sigma^2$:

$$M(\mu_2') = \frac{n-1}{n}\mu_2 \qquad (5_1)$$

Expresia $\sqrt{M(\mu_2')}$ reprezintă media abaterilor tip a selecției. Notînd-o prin $\bar{\sigma}$ avem relația

$$\bar{\sigma} = \sigma\sqrt{\frac{n-1}{n}}. \qquad (5_2)$$

În mod analog se stabilesc și valorile medii

$$M(\mu_3') = \frac{(n-1)(n-2)}{n^2}\mu_3 \qquad (5_3)$$

$$M(\mu_4') = \frac{(n-1)(n^2-3n+3)}{n^3}m_4 + \frac{3(n-1)(2n-3)}{n^3}m_2^3 \qquad (5_4)$$

Să demonstrăm de pildă relația (5_3). Avem din relația (5):

$$M(\mu_3') = M(m_3') - 3M(m_1' \cdot m_2') + 2M(m_1'^3).$$

Dar, $M(m_3') = m_3$ (relația 4_7 pentru $k = 3$)

$$M(m_1' \cdot m_2') = M\left(\frac{\Sigma x_h}{n}\cdot\frac{\Sigma x_j^2}{n}\right) = \frac{1}{n^2}M\left[\sum_h x_h^3 + \sum_{h\neq j} x_h\, x_j^2\right] =$$

$$= \frac{1}{n^2}\sum_{h=1}^{n} M(x_h^3) + \frac{1}{n^2}\sum M(x_h)\, M(x_j^2).$$

Deoarece $M(x_h^3) = M(x^3) = m_3$, pentru orice $h = 1, 2, \ldots n$ avem:

$$\frac{1}{n^2}\sum_{h=1}^{n} M(x_h^3) = \frac{n\,m_3}{n^2} = \frac{m_3}{n}.$$

Suma $\sum_{h\neq j} M(x_h)$, $M(x_j^2) = \sum_{h\neq j} M(x)\cdot M(x^2) = \sum_{h\neq j} m_1 m_2 = n(n-1)\cdot m_1 m_2$

Scanned with OKEN Scanner

întrucît numărul termenilor acestei sume este egal cu n (n − 1).

Aşadar avem :

$$M(m_1' \cdot m_2') = \frac{m_3}{n} + \frac{(n-1)\, m_1\, m_2}{n}.$$

Pentru calculul valorii $M(m_1'^3)$ vom scrie :

$$M(m_1'^3) = M\left(\frac{\Sigma x_h}{n}\right)^3 = \frac{1}{n^3} M(\Sigma x_h)^3 = \frac{1}{n^3} M\left(\sum_h x_h^3 + 3 \sum_{h \neq j} x_h^2 x_j + \right.$$

$$\left. + 6 \sum_{h \neq j \neq l} x_h x_j x_l\right) = \frac{1}{n^3}\left(\sum_h m_3 + 3 \sum_{j \neq h} m_1 m_2 + 6 \sum_{h \neq j \neq l} m_1^3\right).$$

Deoarece $\sum_{j \neq h}$ are n(n − 1) termeni, iar $\sum_{1 \neq j \neq h}$ are $C_n^3 = \frac{n(n-1)(n-2)}{6}$ termeni, obţinem :

$$M(m_1'^3) = \frac{m_3}{n^2} + \frac{3(n-1)\, m_1\, m_2}{n^2} + \frac{(n-1)(n-2)}{n^2} m_1^3.$$

Valoarea medie căutată $M(\mu_3')$ va fi deci :

$$M(\mu_3') = m_3 - \frac{3 m_3}{n} - \frac{3(n-1)\, m_1\, m_2}{n} + \frac{2 m_3}{n^2} + \frac{6(n-1)\, m_1\, m_2}{n^2} +$$

$$+ \frac{2(n-1)(n-2) m_1^3}{n^2} = \frac{(n-1)(n-2)}{n^2} m_3 - 3 \frac{n-1}{n}\left(1 - \frac{2}{n}\right) m_1 m_2 +$$

$$+ \frac{2(n-1)(n-2)}{n^2} m_1^3 = \frac{(n-1)(n-2)}{n^2} (m_3 - 3 m_1 m_2 + 2 m_1^3).$$

Avem însă relaţia :

$$\mu_3 = M(x - m_1)^3 = M(x^3 - 3 x^2 m_1 + 3 x m_1^2 - m_1^3) =$$

$$= M(x^3) - 3 m_1 M(x^2) + 3 m_1^2 M(x) - m_1^3 = m_3 - 3 m_1 m_2 + 2 m_1^3$$

aşa încît relaţia precedentă se scrie :

$$M(\mu_3') = \frac{(n-1)(n-2)}{n^2} \mu_3$$

adică obţinem relaţia (5_3).

Formulele (5_1), (5_3), (5_4) dau distribuțiile exacte ale primelor momente centrate de selecție. Ele sînt îndeosebi utile cînd trebuie să calculăm cu ajutorul lor momentele centrate ale populației Γ.

Observație. Am văzut că dacă avem relațiile :

$$M\,[x\,(n)] = a, \quad \underset{n\to\infty}{M}\,[x\,(n) - a]^2 \to 0 \tag{6}$$

atunci se spune că variabila $x(n)$ depinzînd de n, *estimează absolut corect numărul a.*

Dacă însă :

$$\begin{aligned} &1°. \quad M\,[x\,(n)] = a + A\,(n) \\ &2°. \quad \underset{n\to 0}{M}\,[x\,(n) - a]^2 \to 0 \\ &3°. \quad \underset{n\to\infty}{A\,(n)} \to 0 \end{aligned} \tag{6_1}$$

se spune că $x(n)$ *estimează corect numărul a.* Termenul $A(n)$ se numește *eroare sistematică*; ea descrește o dată cu n.

Exemple : 1°. Dacă de pildă se face o selecție de mărime n dintr-o populație Γ cu distribuție binomială obținem frecvența $\frac{j}{n}$, unde j este numărul care arată de cîte ori s-a realizat evenimentul A care ne interesează, deci A^*, opusul său, s-a realizat de $n - j$ ori. În acest caz, notînd $P(A) = p$, $q = 1 - p$, avem :

$$M\left(\frac{j}{n}\right) = p. \quad M\left(\frac{j}{n} - p\right)^2 = \sqrt{\frac{pq}{n}} \to 0 \text{ pentru } n \to \infty.$$

Rezultă că frecvența $\frac{j}{n}$ *estimează absolut corect probabilitatea p.*

2°. Am văzut mai înainte că :

$$M\,(\mu_2') = \frac{n-1}{n}\,\mu_2 = \mu_2 - \frac{\mu}{n} \quad \text{(relația } 5_1\text{)}$$

$$M\left(\mu_2' - \frac{n-1}{n}\,\mu_2\right)^2 \to 0, \text{ pentru } n \to \infty$$

deci μ_2', *momentul centrat de selecție de ordinul al doilea, estimează numai corect momentul* μ_2 *al populației* Γ. O estimație absolut corectă a lui μ_2 se obține folosind expresia :

$$\frac{n}{n-1}\,\mu_2' = \frac{\sum_{h=1}^{n} (x_h - \bar{x})^2}{n-1}.$$

De aceea, în aplicațiile calculului probabilităților în științele experimentale, estimația momentului centrat de ordinul al doilea se face împărțind suma pătratelor abaterilor față de media de selecție la $(n - 1)$ în loc de n.

În mod analog, μ_3', μ_4' ... sînt estimații corecte pentru momentele centrate μ_3, μ_4, ... ale populației Γ.

§ 3. Selecții de volum redus (selecții mici)

Chestiunile de care ne-am ocupat în paragraful precedent ne dau metode generale pentru aplicarea selecției. Așa de pildă, din faptul că media de selecție estimează absolut corect media colectivității Γ, putem obține o valoare destul de exactă a acesteia din urmă, pe baza unei selecții formată dintr-un număr cît mai mare de elemente. În afară de aceasta, am văzut că media de selecție are o distribuție apropiată de cea normală dacă volumul n al selecției este mare. Aceasta înseamnă că putem obține și unele informații în legătură cu intervalul probabil de variație al valorii estimate. Aceleași concluzii utile se pot trage și asupra momentelor populației Γ de un ordin mai înalt, folosind formulele pe care le-am stabilit mai înainte.

Teoremele demonstrate pînă acum însă au un caracter asimptotic, adică ele dau distribuția limită pentru $n \to \infty$ și ca atare nu pot fi aplicabile selecțiilor de volum redus. În cazuri destul de frecvente, în producție de pildă, este necesar să se tragă concluzii asupra distribuției populației Γ folosind o selecție mică, formată de obicei din cîteva elemente. Asemenea selecții prezintă mari avantaje economisind timp și material, iar în unele cazuri sînt singurele indicate așa după cum am explicat în introducere.

Unor atari selecții, neputînduli-se aplica teoremele generale din § 2, este necesar să li se stabilească distribuția exactă căci numai în acest caz se pot trage concluzii utile din selecția obținută. Este evident că distribuția selecției depinde de distribuția populației Γ și acest lucru complică considerabil calculul determinării distribuției exacte a selecției. În cazul selecției de volum mare am văzut că distribuția mediei de selecție de pildă, era aproximativ normală, oricare ar fi fost distribuția populației Γ: acest rezultat nu mai este adevărat pentru selecțiile de volum redus decît în cazul cînd populația Γ are distribuție normală.

Scanned with OKEN Scanner

În acest paragraf vom expune cîteva probleme care au mari aplicații practice.

1°. *Distribuția binomială.* Să considerăm populația Γ în care am definit variabila aleatorie x care poate lua valorile 1 și 0 cu probabilitățile p și q. Această variabilă are o distribuție binomială și densitatea de probabilitate pentru o singură probă este dată de relația cunoscută :

$$f(x) = p^x \; q^{1-x}, \qquad x = 0,1 \tag{7}$$

Din această populație să facem o selecție de mărime n care ne dă valorile $x_1, x_2, \ldots, x_n$. Această selecție este în realitate un șir de valori 1 sau 0, notînd cu 1 rezultatul probei în care s-a realizat evenimentul A și 0 cînd nu s-a realizat. Media de selecție este :

$$\bar{x} = \frac{1}{n} \sum_{h=1}^{n} x_h \tag{7_1}$$

Am văzut că :

$$M(\bar{x}) = m = p \cdot 1 + 0 \cdot q = p$$

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = \frac{p\,q}{n}$$

Variabila $\bar{x}$ este de tip discret și poate lua numai valorile :

$$0, \; \frac{1}{n}, \; \frac{2}{n}, \; \ldots \; \frac{h}{n}, \ldots 1$$

Funcția de densitate pentru h, adică probabilitatea $P(x = h)$ ca evenimentul A să se întîmple de h ori în n probe este

$$f(h) = C_n^h \; p^h \; q^{n-h}, \qquad h = 0, \; 1, \; 2, \ldots n \tag{7_2}$$

Deoarece $\bar{x} = h/n$, distribuția mediei de selecție $\bar{x}$ va fi

$$g(\bar{x}) = C_n^{n\bar{x}} \; p^{n\bar{x}} \; q^{n(1-\bar{x})}, \quad \bar{x} = 0, \; \frac{1}{n}, \; \frac{2}{n}, \ldots 1. \tag{7_3}$$

Pentru n mare, variabila x are o distribuție normală care se calculează cu formula (4) în care facem $m = p$, $\sigma^2 = pq$.

Altă metodă. Selecția de mărime n din populația Γ cu distribuție binomială, conduce la variabilele aleatorii indepen-

Scanned with OKEN Scanner

dente $x_1, x_2, \ldots x_n$. Densitatea de probabilitate pentru această distribuție n-dimensională va fi :

$$f(x_1, x_2, \ldots, x_n) = \prod_{h=1}^{n} f(x_h) = \prod_{h=1}^{n} p^{x_h} q^{1-x_h}, \quad x_h = 0,1$$

sau încă :

$$f(x_1, x_2, \ldots x_n) = p^{x_1+x_2 \ldots +x_n} q^{n-x_1-x_2-\ldots-x_n}, \quad x_h = 0,1 \qquad (7_4)$$

Media de selecție este :

$$\bar{x} = \frac{1}{n} \sum_{k=1}^{n} x_h, \quad \bar{x} = 0, \frac{1}{n}, \frac{2}{n}, \ldots 1.$$

Densitatea $g\left(\frac{h}{n}\right)$, adică probabilitatea ca $\bar{x}$ să ia valoarea $\frac{h}{n}$, se obține făcînd suma (7_4) pentru toate sistemele de valori :

$$(x_1, x_2, \ldots, x_n) \text{ așa fel ca } \frac{1}{n} \sum_{j=1}^{n} x_j = \frac{h}{n} \text{ adică } \sum_{j=1}^{n} x_j = h.$$

Dar pentru toate aceste sisteme de valori, $f(x_1, x_2, \ldots x_n)$ are aceeași valoare $p^h q^{n-h}$, deci suma se poate calcula înmulțind această valoare cu numărul sistemelor $(x_1, \ldots x_n)$ care îndeplinesc condiția cerută. Numărul acestora este evident C_n^h, deci :

$$g\left(\frac{h}{n}\right) = C_n^h p^h q^{n-h}, \quad \frac{h}{n} = 0, \frac{1}{n}, \frac{2}{n}, \ldots, 1$$

Făcînd $h = n\bar{x}$ obținem densitatea (7_3).

2°. *Distribuția Poisson.* Să găsim distribuția mediei de selecție pentru o populație Γ cu distribuție Poisson. Fie x variabila definită pe Γ; vom avea densitatea :

$$f(x) = e^{-m} \frac{m^x}{x!}, \qquad x = 0, 1, 2, \ldots \qquad (8)$$

unde $M(x) = m$. Media de selecție $\bar{x}$ poate avea evident oricare din valorile $\frac{h}{n}$, $(h = 0, 1, 2, \ldots)$. Pentru valoarea h de pildă, valorile x^j trebuie să satisfacă relația $\frac{h}{n} = \frac{1}{n} \sum x_j$ sau :

$$\sum x_j = h \qquad (8_1)$$

În consecință, ținînd seama că densitatea distribuției n-dimensionale este :

$$f(x_1, x_2, \ldots, x_n) = \prod_{j=1}^{n} f(x_j) = \frac{e^{-nm} m^{\sum_{j=1}^{n} x_j}}{\prod_{j=1}^{n} x_j !} \qquad (8_2)$$

vom putea scrie densitatea :

$$g\left(\frac{h}{n}\right) = \sum_{\Sigma x_j = h} \frac{e^{-nm} m^h}{\Pi x_j !} = e^{-nm} \cdot m^h \sum_{\Sigma x_j = h} \frac{1}{\Pi x_j !}$$

Din formula polinomului însă avem

$$(a_1 + a_2 + \ldots + a_n)^h = \sum \frac{h!}{x_1 ! \ldots x_n !} \prod_{j=1}^{n} a_j^{x_j} = h! \sum \frac{1}{\Pi x_j !} \prod a_j^{x_j}$$

unde $x_1 + x_2 + \ldots + x_n = h$. Făcînd aici $a_1 = a_2 = \ldots = a_n = 1$, găsim :

$$h! \sum_{\Sigma x_j = h} \frac{1}{\Pi x_j !} = n^h$$

De unde rezultă suma căutată :

$$\sum \frac{1}{\Pi x_j !} = \frac{x^h}{h!}$$

iar densitatea $g\left(\frac{h}{n}\right)$ devine :

$$g\left(\frac{h}{n}\right) = \frac{e^{-nm}(nm)^h}{h!}, \quad \frac{h}{n} = \bar{x} = 0, \frac{1}{n}, \frac{2}{n} \ldots 1 \qquad (8_3)$$

sau încă :

$$g(\bar{x}) = \frac{e^{-nm}(nm)^{n\bar{x}}}{(n\bar{x})!} \qquad (8_4)$$

Prin urmare, o selecție de mărime n dintr-o populație cu distribuție Poisson, are tot o distribuție Poisson cu parametrul de n ori parametrul m al distribuției (8).

Scanned with OKEN Scanner

3°. *Distribuția normală.* Populația Γ are o distribuție normală, adică variabila x definită pe această populație are densitatea de probabilitate

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} \tag{9}$$

O selecție de mărime n din această populație m conduce la variabilele aleatorii independente $x_1, x_2, \ldots x_n$ cu densitatea

$$f(x_1, x_2, \ldots, x_n) = \prod_{h=1}^{n} f(x_h) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left(\frac{1}{\sigma^n}\right) e^{-\frac{1}{2}\sum_{h=1}^{n}\left(\frac{x_h - m}{\sigma}\right)^2} \tag{9_1}$$

Făcînd schimbarea de variabile

$$\frac{x_h - m}{\sigma} = \eta_k, \qquad 1 \leqslant h \leqslant n \tag{9_2}$$

obținem densitatea transformată

$$h(\eta_1, \eta_2, \ldots \eta_n) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \cdot \frac{1}{\sigma^n} e^{-\frac{1}{2}\sum_{h=1}^{n}\eta_h^2} \left| \frac{D(x_1\ x_2, \ldots, x_n)}{D(\eta_1, \eta_2, \ldots, \eta_n)} \right|$$

Avem însă :

$$\frac{D(x_1, x_2, \ldots, x_n)}{D(\eta_1, \eta_2, \ldots, \eta_n)} = \begin{vmatrix} \sigma & 0 & \ldots & 0 \\ 0 & \sigma & \ldots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & & \sigma \end{vmatrix} = \sigma^n$$

De unde rezultă :

$$h(\eta_1, \eta_2, \ldots, \eta_n) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} e^{-\frac{1}{2}\sum_{h=1}^{n}\eta_h^2} \tag{9_3}$$

Media de selecție este :

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

De aici obținem :

$$\bar{x} - m = \frac{\Sigma x_h}{n} - m = \frac{\Sigma (x_h - m)}{n} = \frac{\sigma \Sigma \left(\frac{x_h - m}{\sigma} \right)}{n}$$

de unde media :

$$\bar{\eta} = \frac{\Sigma \eta_h}{n} = \frac{\bar{x} - m}{\sigma} \qquad (9_4)$$

Așadar, vom găsi distribuția lui $\bar{\eta}$ și vom deduce apoi distribuția mediei de selecție $\bar{x}$. După metoda generală vom elimina pe η_1 de pildă în (9_4) :

$$\eta_1 = n\bar{\eta} - \sum_2^n \eta_h \qquad (9_5)$$

și obținem densitatea :

$$g(\bar{\eta}, \eta_2, \ldots, \eta_n) = \left(\frac{1}{2\pi} \right)^{\frac{n}{2}} e^{-\frac{1}{2}\left[(n\bar{\eta} - \sum_2^n \eta_h)^2 + \sum_2^n \eta_h^2 \right]} \left| \frac{\partial \eta_1}{\partial \bar{\eta}} \right|$$

Cum $\left| \frac{\partial \eta_1}{\partial \bar{\eta}} \right| = n$, deducem :

$$g(\bar{\eta}, \eta_2, \ldots, \eta_n) = \left(\frac{1}{2\pi} \right)^{\frac{n}{2}} n\, e^{-\frac{1}{2}\left[(n\bar{\eta} - \sum_2^n \eta_h)^2 + \sum_2^n \eta_h^2 \right]} \qquad (9_6)$$

Urmează acum să găsim densitatea marginală în raport cu $\bar{\eta}$, adică să integrăm funcția de mai sus în raport cu variabilele $\eta_2, \eta_3, \ldots \eta_n$ între limitele $-\infty$ și $+\infty$.

În acest scop, considerăm distribuția (9_6) ca o distribuție normală cu variabilele $\bar{\eta}, \eta_2, \ldots \eta_n$. Matricea formei pătratice de la exponent este :

$$\| a_{kh} \| = \begin{Vmatrix} n^2 & -n & -n & \ldots & -n \\ -n & 2 & 1 & \ldots & 1 \\ -n & 1 & 2 & \ldots & 1 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ -n & 1 & 1 & & 2 \end{Vmatrix} \qquad (9_7)$$

Scanned with OKEN Scanner

Determinantul acestei matrici se poate scrie:

$$\left| a_{kh} \right| = n^2 \begin{vmatrix} 1 & -1 & -1 & \dots & -1 \\ -1 & 2 & 1 & \dots & 1 \\ -1 & 1 & 2 & \dots & 1 \\ . & . & . & & . \\ . & . & . & & . \\ . & . & . & & . \\ -1 & 1 & 1 & & 2 \end{vmatrix} = n^2 \begin{vmatrix} 1 & -1 & -1 & \dots & 1 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ . & . & . & & . \\ . & . & . & & . \\ . & . & . & & . \\ 0 & 0 & 0 & \dots & 1 \end{vmatrix} = n^2 \quad (9_8)$$

Primul element al matricei inverse $\| a_{kh} \|^{-1}$ este egal cu complementul algebric al lui n^2 din matricea (9_7) împărțit prin $|a_{kh}| = n^2$, adică:

$$\frac{1}{n^2} \begin{vmatrix} 2 & 1 & \dots & 1 \\ 1 & 2 & \dots & 1 \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ 1 & 1 & \dots & 2 \end{vmatrix} = \frac{1}{n^2} \begin{vmatrix} n & n & \dots & n \\ 1 & 2 & \dots & 1 \\ . & . & & \\ . & . & & \\ . & . & & \\ 1 & 1 & \dots & 2 \end{vmatrix} = \frac{n}{n^2} \begin{vmatrix} 1 & 1 & \dots & 1 \\ 1 & 2 & \dots & 1 \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ 1 & 1 & \dots & 2 \end{vmatrix} =$$

$$= \frac{n}{n^2} \begin{vmatrix} 1 & 1 & \dots & 1 \\ 0 & 1 & \dots & 0 \\ 0 & 0 & 1 & .0 \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ 0 & 0 & 0 & .1 \end{vmatrix} = \frac{n}{n^2} = \frac{1}{n}$$

Inversul acestui element fiind n, rezultă densitatea mediei $\bar{\eta}$:

$$h(\bar{\eta}) = \frac{1}{\sqrt{2\pi}} \cdot \sqrt{n} \quad e^{-\frac{1}{2} n \bar{\eta}^2} \qquad (9_9)$$

Ținînd seama de (9_4), densitatea mediei de selecție $\bar{x}$, va fi:

$$f(\bar{x}) = \frac{1}{\sqrt{2\pi}} \cdot \sqrt{n} \cdot \quad e^{-\frac{1}{2} n \left(\frac{\bar{x}-m}{\sigma}\right)^2} \frac{d\bar{\eta}}{d\bar{x}}$$

și cum $\frac{d\bar{\eta}}{d\bar{x}} = \frac{1}{\sigma}$ obținem:

$$f(\bar{x}) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \quad e^{-\frac{n}{2}\left(\frac{\bar{x}-m}{\sigma}\right)^2} \qquad (9_{10})$$

Am stabilit astfel distribuția mediei de selecție cu parametrii m și $\sigma/\sqrt{n}$ pentru orice populație normală cu dispersia finită. Ea este distribuția oricărei populații cu dispersia finită, însă pentru valori mari ale lui n după cum rezultă din formula (4).

Scanned with OKEN Scanner

Din demonstrația dată însă, rezultă că *media de selecție a unei populații normale cu dispersia finită are o distribuție normală, chiar pentru selecțiile de volum redus.*

Acest rezultat se mai poate deduce și în alt mod.

În acest scop, vom stabili mai întîi următoarea proprietate: Dacă variabilele independente între ele $x_1, x_2, \ldots x_n$, admit fiecare o distribuție normală avînd respectiv mediile $m_1, m_2, \ldots m_n$ și dispersiile $\sigma^2_1, \sigma^2_2, \ldots \sigma^2_n$, atunci variabila aleatorie

$$y_{(n)} = \alpha_1 x_1 + \alpha_2 x_2 + \ldots . \alpha_n x_n \qquad (9_{11})$$

admite o distribuție normală cu media $\sum_{k=1}^{n} \alpha_k \; m_k$ și dispersia

$$\sum_{k=1}^{n} \alpha_k^2 \quad \sigma_k^2.$$

În capitolul VI, §4 formula (18_2) am calculat funcția generatoare a distribuției normale cu media m și dispersia σ^2.

$$m(t) = e^{m\,t + \frac{1}{2} t^2 \sigma^2} \qquad (9_{12})$$

Funcția generatoare a variabilei $y_{(n)}$ va fi:

$$m(t) = M\,(e^{t\,y_{(n)}}) = M\left(e^{t \sum_{k=1}^{n} \alpha_k \; x_k} \right) = M\left(\prod_{k=1}^{n} e^{t\alpha_k \; x_k} \right) =$$

$$= \prod_{k=1}^{n} M\left(e^{t\alpha_k \; x_k} \right)$$

deoarece variabilele x_k sînt independente.

Avem însă, din (9_{12}) înlocuind t prin $t^{\alpha k}$:

$$M\left(e^{t\alpha_k \, x_k} \right) = e^{t\alpha_k \, m_k + \frac{t^2}{2} \alpha_k^2 \; \sigma_k^2}$$

de unde deducem:

$$m(t) = e^{t \sum_{k=1}^{n} \alpha_k m_k + \frac{t^2}{2} \sum_{k=1}^{n} \sigma_k^2 \; \alpha_k^2}$$

Comparînd cu (9_{12}) deducem că variabila $y_{(n)}$ are o distribuție normală cu media $\sum_{k=1}^{n} \alpha_k \, m_k$ și dispersia $\sum_{k=1}^{n} \alpha_k^2 \; \sigma_k^2$.

Scanned with OKEN Scanner

În particular, pentru $\alpha_1 = \alpha_2 = \ldots = \alpha_n = \frac{1}{n}$ obținem media de selecție

$$y_{(n)} = \bar{x} = \frac{x_1 + x_2 + \ldots x_n}{n}$$

unde

$$M(x_1) = M(x_2) = \ldots = M(x_n) = m$$

$$D^2(x_1) = D^2(x_2) = \ldots = D^2(x_n) = \sigma^2$$

Prin urmare, media de selecție $\bar{x}$ are o distribuție normală cu media

$$M(\bar{x}) = \frac{1}{n} \sum_{k=1}^{n} M(x_k) = \frac{m\,n}{n} = m$$

și dispersia

$$D^2(\bar{x}) = \frac{1}{n^2} \sum_{k=1}^{n} \sigma = \frac{1}{n^2}\, n\, \sigma^2 = \frac{\sigma^2}{n}.$$

În aceeași ordine de idei, vom stabili distribuția momentului de selecție de ordinul al doilea a unei populații normale cînd volumul selecției este redus.

Distribuția momentului de selecție de ordinul al doilea. Populația Γ are o distribuție normală cu media 0 și abaterea medie pătratică σ. Să găsim distribuția variabilei aleatorii

$$m_2' = \frac{x_1^2 + x_2^2 + \ldots + x_n^2}{n} \tag{10}$$

Variabilele aleatorii x_k $(1 \leqslant k \leqslant n)$ sînt identice cu variabila x definită pe Γ, deci pentru orice k avem:

$$M(x_k) = 0, \quad D^2(x_k) = \sigma^2 \tag{10_1}$$

Abaterile normate ale variabilelor x_k sînt:

$$\eta_k = \frac{x_k - M(x_k)}{\sigma_k} = \frac{x_k}{\sigma}, \quad 1 \leqslant k \leqslant n \tag{10_2}$$

Momentul de selecție m_2' se poate scrie sub forma:

$$m_2' = \frac{\sigma^2 \sum_{k=1}^{n} \left(\frac{x_k}{\sigma}\right)^2}{n} = \frac{\sigma^2}{n}\, u \tag{10_3}$$

Variabila :

$$u = \sum_{k=1}^{n} \left(\frac{x_k}{\sigma} \right)^2 \qquad (10_4)$$

are o distribuție χ^2 cu n grade de libertate, deci densitatea ei se poate scrie :

$$f(u) = \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} u^{\frac{n}{2}-1} e^{-\frac{1}{2}u} \qquad (10_5)$$

Să găsim acum densitatea funcției de variabila u

$$m_2' = m_2'(u) = \frac{\sigma^2}{n} u \qquad (10_6)$$

Folosind metoda cunoscută, vom rezolva pe u în funcție de m_2', deci :

$$u = \frac{n m_2'}{\sigma^2}, \quad \frac{du}{d m_2'} = \frac{n}{\sigma^2}$$

Densitatea distribuției lui m'_2 va fi deci :

$$g(m_2') = \frac{1}{\Gamma\left(\frac{n}{2}\right) \sigma^{\frac{n}{2}}} \cdot \frac{(n m_2')^{\frac{n}{2}-1}}{(\sigma^2)^{\frac{n}{2}-1}} e^{-\frac{1}{2} \cdot \frac{n m_2'}{\sigma^2}} \frac{n}{\sigma^2}$$

sau încă, efectuînd calculele :

$$g(m_2') = \frac{1}{\Gamma\left(\frac{n}{2}\right)^{\frac{n}{\sigma}}} \left(\frac{n}{2}\right)^{\frac{n}{2}} (m_2')^{\frac{n}{2}-1} e^{-\frac{1}{2}\frac{n m_2'}{\sigma^2}} \qquad (10_7)$$

Probleme

1. Dintr-o populație cu distribuția binomială se extrage selecția x_1, x_2, x_3, x_4. Să se verifice relația :

$$\sum_{x_1, x_2, x_3, x_4} f(x_1, x_2, x_3, x_4) = 1$$

Scanned with OKEN Scanner

2. Să se scrie densitatea mediei pentru o selecție de volum mare cunoscînd media $m = 2,7$ și dispersia $\sigma^2 = 1$ a populației din care s-a extras selecția.

3. Să se demonstreze că expresia $\frac{j}{n}$ estimează corect probabilitatea p a unei distribuții binomiale, unde j este numărul de cîte ori s-a realizat evenimentul A într-o selecție de mărime η.

4. În problema precedentă să se arate că $\frac{j^2}{n^2}$ estimează corect probabilitatea p^2, iar $\frac{j^2 - j}{n\ (n-1)}$ este o estimație absolut corectă.

5. Să se studieze distribuția momentului centrat de selecție de ordinul al doilea pentru selecțiile de volum mic.

Scanned with OKEN Scanner

Capitolul X

TEORIA ESTIMAȚIEI

§ 1. Probleme parametrice

Distribuțiile multora din fenomenele naturale se pot reprezenta printr-o funcție cunoscută care conține unul sau mai mulți parametri. Așa de pildă, dacă fenomenul are o distribuție Poisson, densitatea sa este :

$$f(x;\lambda) = e^{-\lambda}\frac{x^{\lambda}}{x!}, \qquad x = 0,\ 1,\ 2,\ \ldots$$

care conține parametrul λ. Dacă fenomenul are o distribuție normală densitatea sa conține doi parametri, deoarece avem :

$$f(x;\lambda_1,\lambda_2) = \frac{1}{\sqrt{2\pi}\ \lambda_2}\ e^{-\frac{1}{2}\left(\frac{x-\lambda_1}{\lambda_2}\right)^2} \qquad -\infty < x < \infty$$

În aplicațiile calculului probabilităților trebuie să cunoaștem însă distribuția exactă, adică să calculăm valorile parametrilor care intră în structura distribuției. Determinarea acestor parametri poartă numele de *estimarea parametrilor*.

Pentru estimarea parametrilor unei distribuții există mai multe metode. În cele ce urmează vom expune însă numai unele probleme simple în legătură cu aceste estimații.

§ 2. Estimații eficiente

Să presupunem că se cunoaște forma distribuției teoretice a unei variabile aleatorii și că formula care exprimă densitatea de probabilitate (sau funcția de repartiție, deoarece

știm că ambele definesc distribuția statistică) conține unul sau mai mulți parametri $\lambda_1, \lambda_2, \ldots \lambda_n$. Vom considera la început o distribuție unidimensională cu variabila x și un singur parametru λ; densitatea sa o însemnăm prin $f(x; \lambda)$. Stabilirea sau estimarea distribuției după datele selecției este una din problemele de bază ale statisticii matematice, deoarece am văzut că prin cunoașterea densității $f(x)$ sau funcției de repartiție $F(x)$ putem rezolva cele mai importante probleme practice în legătură cu distribuția statistică; în particular, problemele importante privind analiza, compararea și îndeosebi previziunea rezultatelor într-un proces de masă.

În esență, problema estimării parametrului λ cînd forma funcției $f(x; \lambda)$ este cunoscută (legea normală sau legea Poisson etc.) se reduce la determinarea unei funcții de datele $x_1, x_2, \ldots x_n$ ale unei selecții de mărime n, adică:

$$\lambda^* = \lambda^* (x_1, x_2, \ldots x_n) \tag{1}$$

care pentru diferite selecții să aibă valorile concentrate în jurul parametrului necunoscut λ. Din punct de vedere practic putem deci scrie:

$$\lambda^* (x_1, x_2, \ldots x_n) \approx \lambda \tag{1_1}$$

Problema estimării unui număr de r parametrii ai distribuției $f(x; \lambda_1, \lambda_2, \ldots \lambda_r)$ se pune în același mod, adică să găsim funcțiile $\lambda_1^* (x_1 \ldots x_n), \ldots \lambda_r^* (x_1 \ldots x_n)$ care pentru diferite selecții să verifice relațiile aproximative:

$$\lambda_k^* (x_1, x_2 \ldots x_r) \approx \lambda_k, \quad 1 \leqslant k \leqslant r$$

adică valorile funcțiilor λ_k^* să fie concentrate în jurul parametrului estimat $\lambda_k (1 \leqslant k \leqslant r)$.

Pentru simplificarea expunerii să revenim la cazul unui singur parametru λ.

Am văzut că λ^* estimează absolut corect pe λ dacă avem îndeplinite condițiile:

$$M [\lambda^* (x_1, x_2 \ldots x_n)] = \lambda$$

$$D^2 (\lambda^*) = M (\lambda^* - \lambda)^2 \underset{n \to \infty}{\to} 0 \tag{1_2}$$

Dacă însă $M [\lambda^*] = \lambda + \varphi(n)$, unde $\varphi(n) \to 0$ dacă $n \to \infty$ și dispersia $M (\lambda^* \underset{n \to \infty}{\to} \lambda)^2 \to 0$, atunci λ^* estimează corect pe λ.

Estimații absolut corecte ale unui parametru se obțin în diverse moduri. Așa de pildă, valoarea medie a unei populații

normale poate fi estimată fie prin media de selecție $\bar{x}$, fie prin mediana selecției. Dintre două estimații absolut corecte ale parametrului λ fie ele $\lambda^*(x_1, x_2, \ldots x_n)$, $\lambda^{**}(x_1, x_2, \ldots x_n)$ vom considera mai bună pe aceea care are dispersia mai mică în jurul parametrului estimat. Măsura dispersiei este dată de relația :

$$D^2(\lambda^*) = \sigma^2_{\lambda^*} = M\left[\lambda^*(x_1, x_2, \ldots x_n) - \lambda\right]^2 \qquad (1_3)$$

Așadar, se pune problema găsirii *unei estimații absolut corecte cu dispersia minimă*. În acest scop vom demonstra că dispersia estimațiilor obținute printr-o selecție de mărime n admite o anumită margine inferioară și anume :

$$D^2(\lambda^*) = \sigma^2_{\lambda^*} \geqslant \frac{1}{n.M\left\{\left[\frac{\partial \ln f(x;\lambda)}{\partial \lambda}\right]^2\right\}} \qquad (1_4)$$

În adevăr $f(x;\lambda)$ fiind o densitate de probabilitate verifică relația :

$$\int_{-\infty}^{+\infty} f(x;\lambda)\,dx = 1 \qquad (1_5)$$

de unde, derivînd în raport cu λ deducem :

$$\frac{d}{d\lambda}\int_{-8}^{+\infty} f(x;\lambda)\,dx = \int_{-\infty}^{+\infty} \frac{\partial f(x;\lambda)}{\partial \lambda}\,dx = \int_{-\infty}^{+\infty} \frac{\ln f(x;\lambda)}{\partial \lambda} f(x;\lambda)\,d\lambda = 0 \qquad (1_6)$$

Se știe că valoarea medie a funcției $\varphi(x)$ unde variabila aleatorie x are densitatea $f(x;\lambda)$ este egală cu $\int_{-\infty}^{+\infty} \varphi(x) f(x;\lambda)dx$, deci ultima integrală din relația precedentă este evident

$$M\left[\frac{\partial \ln f(x;\lambda)}{\partial \lambda}\right] = 0 \qquad (1_7)$$

Pe de altă parte avem valoarea probabilă

$$M\left[\lambda^*(x_1, x_2, \ldots x_n)\right] =$$

$$= \int\int \ldots \int_{(R)} \lambda^*(x_1, x_2, \ldots x_n) f(x_1;\lambda) \ldots f(x_n;\lambda)\,dx_1 \ldots dx_n \qquad (1_8)$$

unde R reprezintă domeniul în care se găsește punctul aleator $P(x_1, x_2, \ldots x_n)$.

Scanned with OKEN Scanner

Derivînd în raport cu λ și ținînd seama de relația (1_7) găsim :

$$\int\int \dots \int_{(R)} (\lambda^* - \lambda) \sum_{k=1}^{n} \frac{\partial \ln f(x_k ; \lambda)}{\partial \lambda} \prod_{k=1}^{n} f(x_k ; \lambda) \prod_{k=1}^{n} dx_k = 1 \qquad (1_9)$$

din care deducem :

$$M\left[(\lambda^* - \lambda) \sum_{k=1}^{a} \frac{\partial \ln f(x_k ; \lambda)}{\partial \lambda} \right] = 1 \qquad (1'_9)$$

În baza inegalității lui Schwartz[1] putem scrie :

$$\left\{ M\left[(\lambda^* - \lambda) \sum_{k=1}^{a} \frac{\partial \ln f(x_k ; \lambda)}{\partial \lambda} \right] \right\}^2 \leq M(\lambda^* - \lambda)^2 \cdot M\left[\sum_{k=1}^{a} \frac{\partial \ln f(x_k ; \lambda)}{\partial \lambda} \right]^2$$

Ținînd seama de relațiile (1_7) și $(1'_9)$ găsim :

$$D^2(\lambda^*)\, D^2\left[\sum_{k=1}^{n} \frac{\partial \ln f(x_k ; \lambda)}{\partial \lambda} \right] \geq 1$$

Deoarece variabilele aleatorii $x_1, x_2 \dots x_n$ sînt independente între ele și toate identice cu x putem scrie relația :

$$D^2\left[\sum_{k=1}^{n} \frac{\partial \ln f(x_k ; \lambda)}{\partial \lambda} \right] = \eta D^2\left[\frac{\partial \ln f(x ; \lambda)}{\partial \lambda} \right]$$

Înlocuind în relația de mai sus obținem relația (1_4).

Mai departe vom demonstra relația :

$$M\left[\frac{\partial \ln f(x ; \lambda)}{\partial \lambda} \right]^2 = -M\left[\frac{\partial^2 \ln f(x ; \lambda)}{\partial \lambda^2} \right] \qquad (1_{10})$$

așa încît, inegalitatea (1_4) se mai poate scrie :

$$D^2(\lambda^*) = \sigma^2_{\lambda^*} \geq \frac{1}{-nM\left[\frac{\partial^2 \ln f(x ; \lambda)}{\partial \lambda^2} \right]} \qquad (1'_4)$$

---

[1] Dacă f(x) și $\varphi(x)$ sînt integrabile în intervalul (a, b), subzistă inegalitatea

$$\left[\int_a^b f(x)\, \varphi(x)\, dx \right]^2 \leq \int_a^b f^2(x)\, dx \int_a^b \varphi^2(x)\, dx$$

Scanned with OKEN Scanner

D e f i n i ț i e. Estimația $\bar{\lambda}^* = \bar{\lambda}^* (x_1, x_2, \ldots x_n)$ pentru care dispersia $D^2(\bar{\lambda}^*)$ este egală cu marginea inferioară.

$$\frac{1}{n\,M\left[\frac{\partial \ln f(x;\lambda)}{\partial \lambda}\right]^2} = \frac{1}{-\,nM\left[\frac{\partial^2 \ln f(x;\lambda)}{\partial \lambda^2}\right]} \qquad (1'_5)$$

se numește *estimație eficientă*, iar raportul

$$e(\lambda^*) = \frac{\sigma^2_{\lambda^*}}{\sigma^2_{\lambda^*}} \leq 1 \qquad (1_{11})$$

este *raport de eficiență*.

Exemplu. Să considerăm distribuția Poisson

$$f(x;\lambda) = \sigma^{-\lambda} \cdot \frac{\lambda^*}{x!},\ x = 0,\ 1,\ 2,\ \ldots \qquad (1_{12})$$

și să verificăm dacă media de selecție este o estimație absolut corectă și eficientă a parametrului λ. Avem succesiv

$$\ln f(x;\lambda) = -\lambda + x \ln \lambda - \ln x!$$

$$\frac{\partial \ln f(x;\lambda)}{\partial \lambda} = \frac{x}{\lambda} - 1,\quad \frac{\partial^2 \ln f(x;\lambda)}{\partial \lambda^2} = -\frac{x}{\lambda^2}$$

De unde marginea inferioară:

$$\frac{1}{-\,nM\left[\frac{\partial^2 \ln f(x;\lambda)}{\partial \lambda^2}\right]} = \frac{\lambda^2}{nM(x)} = \frac{\lambda}{n}$$

deoarece pentru distribuția (1_{12}) avem $M(x) = \lambda$.

Media aritmetică

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

are valoarea medie:

$$M(\bar{x}) = \frac{\sum_{k=1}^{a} M(x_k)}{n} = \frac{nM(x)}{n} = \frac{n\lambda}{n} = \lambda$$

întrucît variabilele aleatorii $x_1, x_2 \ldots x_n$ au aceeași distribuție ca și x. Dispersia sa este

$$D^2(\bar{x}) = \frac{1}{n^2} \sum_{k=1}^{n} D^2(x_k) = \frac{n\,D^2(x)}{n^2} = \frac{D^2(x)}{n} = \frac{\lambda}{n}$$

Scanned with OKEN Scanner

întrucît pentru distribuția Poisson am văzut că $D^2(x) = \lambda$ Dispersia mediei de selecție $\bar{x}$ fiind egală cu marginea inferioară găsită mai sus rezultă că $\bar{x}$ este o estimație eficientă pentru parametrul λ.

Alt exemplu. Să considerăm o populație Γ în care o parte din elementele sale posedă caracteristica A și o parte nu o posedă. Făcînd o selecție de mărime n după schema bilei revenite (selecție repetată) s-au obținut valorile $x_1, x_2, \ldots x_n$, unde evident variabilele x_k $(1 \leqslant k \leqslant n)$ iau valorile 1 sau 0 cu probabilitățile p și $q = 1 - p$, după cum în proba n s-a realizat sau nu evenimentul A. Să estimăm probabilitatea necunoscută p cu ajutorul frecvenței relative.

Notînd cu x de cîte ori s-a realizat A în cele n probe ale selecției, avem densitatea cunoscută :

$$f(x\,;p) = C_n^x p^x (1-p)^{n-x},\ x = 0,\ 1,\ 2. \ldots n \qquad (1_{13})$$

De unde :

$$\ln f(x\,;p) = \ln C_n^x + x \ln p + (n-x) \ln (1-p)$$

$$\frac{\ln f(x\,;p)}{p} = \frac{x}{p} - \frac{n-x}{1-p}$$

$$\frac{2 \ln f(x\,;p)}{p^2} = -\frac{x}{p^2} = \frac{n-x}{(1-p)^2}$$

$$M\left\{\frac{\partial^2 \ln f(x\,;p)}{\partial p^2}\right\} = -\frac{M(x)}{p^2} + \frac{n - M(x)}{(1-p)^2} = -\frac{np}{p^2} + \frac{n(1-p)}{(1-p)^2} =$$

$$= -n\left(\frac{1}{p} + \frac{1}{1-p}\right) = -\frac{n}{pq}$$

deoarece pentru distribuția binomială $M(x) = np$.
Rezultă de aici marginea inferioară a dispersiilor

$$\sigma_p^2 \geqq \frac{1}{(-n)\ (-n)\ \dfrac{1}{pq}} = \frac{pq}{n^2}$$

Să considerăm acum media (frecvența relativă) cu care să estimăm parametrul p. Vom avea :

$$p^*(x_1, x_2 \ldots x_n) = \frac{1}{n} \sum_{k=1}^{n} x_k = \bar{x}$$

Scanned with OKEN Scanner

pentru care sînt cunoscute valorile :

$$M(p^*) = p, \; D^2(p) = \frac{p\,q}{n}$$

Această estimație absolut corectă nu este și eficientă, deoarece $\frac{p\,q}{n} > \frac{p\,q}{n^2}$ marginea găsită mai sus.

§ 3. Estimații suficiente

Să considerăm densitatea de probabilitate $f(x;\lambda)$ depinzînd de un singur parametru λ și să efectuăm o selecție de mărime n care dă valorile $x_1, x_2 \ldots x_n$. Densitatea asociată distribuției n—dimensionale este dată de funcția $P_n(x_1, x_2, \ldots x_n;\lambda)$ numită *funcție de verosimilitate* și care este egală cu produsul :

$$P_n(x_1, x_2 \ldots x_n) = \prod_{k=1}^{n} f(x_k;\lambda) \tag{2}$$

deoarece variabilele $x_1, x_2 \ldots x_n$ sînt independente.

Dacă funcția P_n se poate pune sub forma unui produs de doi factori în modul următor :

$$P_n = f_1[\lambda^*(x_1, x_2 \ldots x_n);\lambda]\, f_2(x_1, x_2, \ldots x_n;\lambda) \tag{2_1}$$

în care factorul f_1 este densitatea estimației λ^* pentru o anumită valoare dată a parametrului λ, iar f_2 reprezintă densitatea condiționată a sistemului de n variabile $x_1, x_2 \ldots x_n$ pentru o valoare fixă a parametrului λ^*, atunci $\lambda^*(x_1, x_2, \ldots x_n)$ se numește *estimație suficientă a parametrului* λ.

Exemplu. Să considerăm distribuția Poisson :

$$f(x;\lambda) = e^{-\lambda}\,\frac{\lambda^x}{x!}, x = 0, 1, 2 \ldots \tag{2_2}$$

Densitatea asociată selecției $x_1, x_2 \ldots x_n$ va fi :

$$P_n = e^{-n\lambda}\,\frac{\lambda^{\Sigma x_k}}{\prod_{k=1}^{n} x_k!} \tag{2_3}$$

Să considerăm ca estimație λ^* media de selecție, adică :

$$\lambda^*(x_1, x_2 \ldots x_n) = \frac{\Sigma x_k}{n} = \bar{x} \tag{2_4}$$

Probabilitatea P_n se poate scrie sub forma

$$P_n = \frac{e^{-n\lambda}\lambda^{n\lambda^*}}{(n\lambda^*)!} \cdot \frac{(n\lambda^*)!}{\pi x_k!} \qquad (2_5)$$

Primul factor $e^{-n\lambda}\lambda^{n\lambda^*}/(n\lambda^*)!$ reprezintă probabilitatea egalității $\lambda^*(x_1, x_2 \ldots x_n) = \frac{\sum x_k}{n}$! sau, ceea ce este același lucru, ca $\sum_{k=1}^{n} x_k = n\lambda^*$.

În adevăr, să presupunem că variabilele x_k $(1 \leqslant k \leqslant n)$ urmează legea Poisson, respectiv cu parametrii λ. Avem densitățile

$$f(x_k; \lambda_k) = e^{-\lambda_k} \frac{\lambda_k^{x_k}}{x_k!}, \qquad x_k = 0, 1, 2 \ldots \qquad (2_6)$$

Funcția generatoare a variabilei x_k este dată de relația cunoscută :

$$m_k(t) = e^{\lambda k (e^t - 1)}, \qquad k = 1, 2 \ldots n$$

iar funcția generatoare a sumei $\sum_{k=1}^{n} x_k$ este egală cu produsul funcțiilor generatoare ale termenilor, adică :

$$m(t) = \prod_{k=1}^{n} m_k(t) = e^{(e^t - 1)\sum_{k=1}^{n} \lambda_k} \qquad (2_8)$$

Am găsit astfel funcția generatoare tot a unei distribuții Poisson cu parametrul $\lambda_1 + \lambda_2 + \ldots + \lambda_n$ egal cu suma parametrilor celor n distribuții. În consecință suma variabilelor $x_1, x_2 \ldots x_n$, identice cu x, va avea tot o distribuție Poisson cu parametrul $n\lambda$.

Al doilea factor din (2_1) reprezintă probabilitatea condiționată ca selecția să dea valorile $x_1, x_2 \ldots x_n$, în ipoteza că $\Sigma x_k = n\lambda^*$. Aceasta rezultă din distribuția multinomială. Așadar, $\lambda^*(x_1 \ldots x) = \bar{x}$ este o estimație suficientă pentru λ.

Pentru r parametrii $\lambda_1, \lambda_2 \ldots \lambda_r$ estimația suficientă se definește în mod analog. Astfel sistemul

$$\lambda_k^*(x_1, x_2, \ldots x_n), \quad 1 \leq k \leq r \qquad (2_9)$$

Scanned with OKEN Scanner

se numește *sistem de estimații suficiente* pentru parametrii $\lambda_1, \lambda_2, \ldots \lambda_r$, dacă densitatea asociată datelor selecției se poate pune sub forma :

$$P_n = f_1 \left[\lambda_1^* (x_1, \ldots x_n), \ldots \lambda_r^* (x_1, \ldots x_r) ; \lambda_1 \ldots \lambda_r \right] \cdot$$
$$f_2 (x_1, x_2 \ldots x_n ; \lambda_1^*, \lambda_2^*, \ldots \lambda_r^*) \qquad (2_{10})$$

Exemplu. Fie distribuția normală

$$f (x ; \lambda_1, \lambda_2) = \frac{1}{\sqrt{2\pi\lambda_2}} e^{-\frac{1}{2\lambda_2} (x-\lambda_1)^2} \qquad (2_{11})$$

Densitatea asociată datelor selecției este :

$$P_n = \left(\frac{1}{2\pi} \right)^{\frac{n}{2}} \sigma^{-n} e^{-\frac{1}{2\lambda_2} \sum_{k=1}^{n} (x_k - \lambda_1)^2} \qquad (2_{12})$$

Să considerăm estimația

$$\lambda_1^* (x_1 x_2 \ldots x_n) = \frac{x_1 + x_2 + \ldots + x_n}{n} = \bar{x} \qquad (2_{13})$$

și să scriem suma de la exponent sub forma :

$$\sum (x_k - \lambda_1)^2 = \sum (x_k - \lambda_1^*, + \lambda_1^* - \lambda_1)^2 = \sum (x_k - \lambda_1^*)^2 +$$
$$+ 2 (\lambda_1^* - \lambda_1) \sum (\lambda_k - \lambda_1^*) + n \sum (\lambda_1^* - \lambda)^2$$

În acest caz, avînd :

$$\sum (x_k - \lambda_1^*) = \sum x_k - n \lambda_1^* = 0$$

vom obține :

$$\sum (x_k - \lambda_1)^2 = \sum (x_k - \lambda_1^*)^2 + n (\lambda_1^* - \lambda)^2 \qquad (2_{14})$$

Dacă introducem încă notația

$$\lambda_2^* = \bar{\sigma}^2 = \frac{1}{n} \sum (x_k - \lambda_1^*)^2 \qquad (2_{15})$$

putem scrie :

$$P_n = \left(\frac{1}{2\pi} \right)^{\frac{n}{2}} \bar{\sigma}^n e^{-\frac{n\lambda_1^*}{2\lambda_2}} e^{-\frac{n \cdot (\lambda_1^* - \lambda_1)^2}{2\lambda_2^2}} \qquad (2_{16})$$

adică P_n este de forma (2_{10}) unde $f_2 = 1$

Scanned with OKEN Scanner

Prin urmare, $\lambda_1^* = x$ și $\lambda_2^* = \overline{\sigma}^2$ sînt estimații suficiente pentru parametrii λ_1 și λ_2 ai legii normale. Estimațiile suficiente conțin, după cum rezultă din definiția lor, aproape toate informațiile despre parametrii estimați ; ele sînt estimațiile cele mai bune ale parametrilor.

§ 4. Metoda momentelor

În populația Γ s-a definit variabila aleatorie x cu densitatea de probabilitate $f(x; \lambda_1, \lambda_2, .. \lambda_r)$ depinzînd de r parametrii $\lambda_1, \lambda_2, \ldots \lambda_r$. Se știe că funcția

$$\lambda^* = \lambda^* (x_1, x_2, .. x_n) \tag{3}$$

estimează absolut corect parametrul λ dacă :

$$1^\circ \; M(\lambda^*) = \lambda$$
$$2^\circ \; D^2(\lambda) = M(\lambda^* - \lambda)^2 = 0 \tag{3_1}$$
$$n \to \infty \qquad n \to \infty$$

Să extragem din populația Γ o selecție de mărime n obținînd valorile de selecție $x_1, x_2, \ldots x_n$, unde n este suficient de mare. Am calculat mai înainte valorile probabile ale momentului de selecție

$$m_k' = \frac{x_1^k + x_2^k + \ldots + x_n^k}{n} = m_k'(x_1 x_2, \ldots x_n) \tag{3_2}$$

și anume pentru orice $k = 1, 2, \ldots$ avem relațiile :

$$M(m_k') = m_k, \; M(m_k' - m_k)^2 = D^2(m_k') = \frac{m_{2k} - m_k}{n} \tag{3_3}$$

De aici rezultă că momentele de selecție estimează absolut corect momentele teoretice ale populației $m_1, m_2, \ldots$

Metoda momentelor, imaginată de K. Pearson, constă în esență din egalarea primelor r momente teoretice ale distribuției cu densitatea $f(x; \lambda_1, \lambda_2 .. \lambda_r)$ cu primele r momente empirice și rezolvarea ecuațiilor obținute în raport cu cei r parametrii necunoscuți $\lambda_1, \lambda_2, \ldots \lambda_r$.

În adevăr, din relația

$$m_k = \int_{-\infty}^{+\infty} x^k f(x; \lambda_1, \lambda_2, \ldots \lambda_r) \, dx$$

obținem funcția

$$m_k = m_k(\lambda_1, \lambda_2, \ldots \lambda_r), \quad 1 \leq k \leq r \tag{3_4}$$

De unde deducem rezolvînd în raport cu $\lambda_1, \lambda_2, \ldots \lambda_r$ sistemul precedent de r ecuații :

$$\lambda_k = \lambda_k(m_1, m_2, \ldots m_r), \quad 1 \leqslant k \leqslant r \tag{3_5}$$

Dacă în locul momentelor teoretice necunoscute $m_1, m_2, \ldots m_r$ se introduc valorile lor aproximative date de selecție $m_1', m_2', \ldots m_r'$ se obțin estimațiile

$$\lambda_k^* = \lambda\left[m_1'(x_1, x_2, \ldots x_n), \ldots m_r'(x_1, x_2, \ldots x_n)\right] = \lambda^*(x_1, x_2, \ldots x_n) \tag{3_6}$$

Exemplu. Fie distribuția depinzînd de parametrul λ

$$f(x;\ \lambda) = \frac{1}{\Gamma(\lambda)} x^{\lambda-1} e^{-x}, \quad x > 0,\ \lambda > 0$$

Se face o selecție de mărime n din care se obțin valorile $x = x_{12}\ x_2 \ldots x_n$ și se cere să se estimeze parametrul λ.

Momentul de ordinul întîi al distribuției este :

$$m = \int_{-\infty}^{+\infty} x\, f(x;\ \lambda)\, d_x = \int_0^{\infty} \frac{x^{\lambda-1} e^{-x}}{\Gamma(\lambda)}\, dx = \frac{\Gamma(\lambda+1)}{\Gamma(\lambda)} = \lambda.$$

Din egalarea acestuia cu momentul de selecție obținem :

$$m_1' = x = \lambda^*(x_1, x_2, \ldots x_n) = \frac{1}{n}(x_1 + x_2 + \ldots + x_n).$$

Se vede imediat că $M(\lambda^*) = \frac{1}{n}\sum M(x_k) = \frac{n\lambda}{n} = \lambda.$

§ 5. Metoda verosimilității maxime

O metodă de estimare care va juca un rol deosebit în cele ce urmează este așa numita metodă a *verosimilității maxime* a lui R.A. Fisher pe care o enunțăm în modul următor :

Se consideră o populație Γ în care s-au definit s variabile aleatorii $x_1, x_2, \ldots, x_s$; densitatea de probabilitate a acestei distribuții s-dimensionale depinde de r parametrii $\lambda_1, \lambda_2, \ldots \lambda_r$ deci este o funcție de forma $f(x_1, x_2, \ldots x_s;\ \lambda_1, \lambda_2, \ldots \lambda_r)$ pe

care adesea o vom scrie mai scurt $f(x; \lambda_1, \lambda_2, \ldots \lambda_r)$. Într-o selecție de mărime n, variabila x_1 poate lua valorile

$$x_{11}, x_{12}, \ldots, x_{1n}$$

variabila x_2 poate lua valorile

$$x_{21}, x_{22}, \ldots, x_{2n}$$

. . . . . .

variabila x_s poate lua de asemenea n valori

$$x_{s1}, x_{s2}, \ldots x_{sn}.$$

Densitatea de probabilitate asociată datelor experienței este funcția de verosimilitate :

$$P_n = \prod_{k=1}^{n} f(x_k; \lambda_1, \lambda_2, \ldots \lambda_r) \tag{4}$$

unde prin x_k am notat ansamblul valorilor luate de cele s variabile în proba k.

Valorile parametrilor $\lambda_1, \lambda_2, \ldots, \lambda_r$ se determină cu condiția ca probabilitatea P_n să aibă valoarea maximă. Deoarece maximul lui P_n are loc o dată cu maximul lui $\ln P_n$, unde ln însemnează logaritmul natural, notînd $L = \ln P_n$, valorile parametrilor λ_k $(1 \leqslant k \leqslant r)$ se determină din sistemul de r ecuații cu r necunoscute :

$$\frac{\partial L}{\partial \lambda_1} = 0, \quad \frac{\partial L}{\partial \lambda_2} = 0, \ldots \frac{\partial L}{\partial \lambda_r} = 0 \tag{4_1}$$

Soluțiile acestui sistem pe care le notăm prin $\lambda_1^*, \lambda_2^*, \ldots, \lambda_r^*$ se numesc *indicatorii de probabilitate maximă* și ei reprezintă estimațiile parametrilor $\lambda_1^*, \lambda_2^*, \ldots \lambda_r^*$.

Aplicație. Se consideră o populație normală bidimensională Γ; această populație se poate considera formată din elemente care posedă două caracteristici A_1 și A_2. Așa de exemplu rezultatele unei serii de trageri într-o țintă sînt caracterizate prin abscisele punctelor în care ținta a fost lovită (caracteristica A_1) și prin ordonatele acelorași puncte (caracteristica A_2). Piesele produse de o mașină pot fi caracterizate de pildă prin mărimea unei anumite dimensiuni (A_1) și prin mărimea unui anumit unghi (A_2).

Așadar, unui element oarecare al populației îi corespunde o pereche de numere (x_1, x_2) sau, în interpretarea geometrică pe care am folosit-o adesea, unui element al populației Γ îi cores-

punde punctul de coordonate (x_1, x_2) față de un sistem $x_1\ 0x_2$ de axe rectangulare. Populația este formată din mulțimea punctelor din planul $x_1\ 0x_2$, care corespund tuturor sistemelor de valori posibile date variabilelor x_1 și x_2. Distribuția acestei populații va reprezenta o suprafață în spațiul cu trei dimensiuni, un punct oarecare al suprafeței avînd coordonatele $[x_1, x_2, f(x_1, x_2)]$, unde $f(x_1, x_2)$ este densitatea de probabilitate a distribuției bidimensionale.

Întrucît Γ este o populație normală, densitatea sa este :

$$f(x_1, x_2) = \frac{\sqrt{|(a_{kh})|}}{2\pi} e^{-\frac{1}{2}\sum_{k=1}^{2}\sum_{h=1}^{2} a_{kh}(x_k - m_k)(x_h - m_h)} \tag{4_2}$$

Forma pătratică de la exponent se poate scrie dezvoltat :

$$a_{11}(x_1 - m_1)^2 + 2a_{12}(x_1 - m_1)(x_2 - m_2) + a_{22}(x_2 - m_2)^2$$

deci ea conține 5 parametri necunoscuți :

$$m_1,\ m_2;\ a_{11},\ a_{12},\ a_{22} \tag{4_3}$$

Ne propunem să estimăm acești parametri efectuînd o selecție de mărime n din populația Γ, pe care o presupunem suficient de numeroasă pentru a nu mai ține seama că selecția este repetată sau nerepetată.

Selecția întîmplătoare din Γ ne-a dat rezultatele :

$$(x_{11}, x_{21}),\ (x_{12}, x_{22}),\ (x_{13}, x_{23}) \ldots (x_{1n}, x_{2n}) \tag{4_4}$$

sau mai scurt :

$$(x_{1i}, x_{2i}),\quad i = 1, 2, \ldots, n \tag{4_4'}$$

Am obținut astfel 2 n variabile aleatorii

$$\begin{matrix} x_{11}, x_{12}, \ldots x_{1i} \ldots x_{1n} \\ x_{21}, x_{22}, \ldots x_{2i} \ldots x_{2n} \end{matrix} \tag{4_4''}$$

din care primele n sînt identice cu x_1, iar celelalte sînt identice cu x_2 căci înainte de efectuarea selecției, fiecare variabilă x_{1i} de pildă, poate lua orice valoare ca și x_1.

Vom putea deci scrie relațiile :

$$\begin{aligned} M(x_{1i}) &= M(x_1) = m_1 \\ M(x_{2i}) &= M(x_2) = m_2 \\ D^2(x_{1i}) &= D^2(x_1) = \sigma_1^2 \\ D^2(x_{2i}) &= D^2(x_2) = \sigma_2^2 \end{aligned} \tag{4_5}$$

pentru orice $i = 1, 2, \ldots n$.

Scanned with OKEN Scanner

Densitatea variabilelor (x_{1i}, x_{2i}), după relația (2_2) va fi :

$$f_i(x_{1i}, x_{2i}) = \frac{\sqrt{|a_{kh}|}}{2\pi} e^{-\frac{1}{2}\sum_{k=1}^{2}\sum_{h=1}^{2} a_{kh}(x_{ki}-m_k)(x_{hi}-m_h)} \qquad (4'_2)$$

Pentru întreaga selecție, în care extracțiile sînt independente între ele, vom avea densitatea

$$\bar{f} = \prod_{i=1}^{n} \bar{f}_i(x_{1i}, x_{2i}) = \frac{|a_{kh}|^{\frac{n}{2}}}{(2\pi)^n} e^{-\frac{1}{2}\sum_{k=1}^{2}\sum_{h=1}^{2}\sum_{i=1}^{2} a_{kh}(x_{ki}-m_k)(x_{hi}-m_h)} \qquad (2''_2)$$

Pentru estimarea celor 5 parametri relația (2_3), derivăm funcția $L = \ln \bar{f}$, în raport cu acești parametri, și rezolvăm ecuațiile astfel obținute.

Avem logaritmul :

$$L = -n \ln 2\pi + \frac{1}{2} n \ln (a_{11} a_{22} - a_{12}^2) - \frac{1}{2}\sum_{i=1}^{n} [a_{11}(x_{1i}-m_1)^2 +$$

$$+ 2 a_{12}(x_{1i}-m_1)(x_{2i}-m_2) + a_{22}(x_{2i}-m_2)^2] \qquad (4_6)$$

De unde rezultă derivatele parțiale :

$$\frac{\partial L}{\partial m_1} = -\sum_i a_{11}(x_{1i}-m_1) - \sum_i a_{12}(x_{2i}-m_2)$$

$$\frac{\partial L}{\partial m_2} = -\sum_i a_{22}(x_{2i}-m_2) + \sum_i a_{12}(x_{1i}-m_1)$$

$$\frac{\partial L}{\partial a_{11}} = \frac{n}{2}\frac{a_{22}}{|a_{kh}|} - \frac{1}{2}\sum_i (x_{1i}-m_1)^2$$

$$\frac{\partial L}{\partial a_{12}} = -n\frac{a_{12}}{|a_{kh}|} - \sum_i (x_{1i}-m_1)(x_{2i}-m_2)$$

$$\frac{\partial L}{\partial a_{22}} = \frac{n}{2}\cdot\frac{a_{11}}{|a_{kh}|} - \frac{1}{2}\sum_i (x_{2i}-m_2)^2$$

Scanned with OKEN Scanner

Vom avea deci sistemul :

$$a_{11} \sum_i (x_{1i} - m_1) + a_{12} \sum_i (x_{2i} - m_2) = 0$$

$$a_{12} \sum_i (x_{1i} - m_1) + a_{22} \sum_i (x_{2i} - m_2) = 0 \qquad (4_7)$$

$$a_{11} = |a_{kh}| \frac{\sum_i (x_{2i} - m_2)^2}{n}$$

$$a_{12} = -|a_{kh}| \frac{\sum_i (x_{1i} - m_1)(x_{2i} - m_2)}{n}$$

$$a_{22} = |a_{kh}| \frac{\sum_i (x_{1i} - m_1)^2}{n}$$

Din primele două ecuații în care considerăm ca necunoscute $\sum_i (x_{1i} - m_1)$ și $\sum_i (x_{2i} - m_2)$ obținem :

$$\sum_i (x_{1i} - m_1) = 0, \qquad \sum_i (x_{2i} - m_2) = 0$$

deoarece determinantul sistemului (care este omogen) este $|a_{kh}| = a_{11}\, a_{22} - a_{12}^2 \neq 0$. Din aceste relații deducem estimațiile

$$m_1^* = \frac{1}{n} \sum_{i=1}^{n} x_{1i}$$

$$m_2^* = \frac{1}{n} \sum_{i=1}^{n} x_{2i} \qquad (4_8)$$

Prin urmare, parametrii m_1 și m_2 din densitatea populației Γ dată de (2_2) se pot estima cu ajutorul datelor selecției și sînt precis mediile acestor date.

Să trecem acum la estimarea ultimilor trei parametri a_{11}, a_{12}, a_{22}. În acest scop, reamintim că dispersiile datelor de selecție pentru valorile x_{1i}, respectiv x_{2i} sînt :

$$\bar{\sigma}_1^2 = \frac{\sum_i (x_{1i} - m_1^*)}{n}, \qquad \bar{\sigma}_2^2 = \frac{\sum_i (x_{2i} - m_2^*)^2}{n} \qquad (4_9)$$

Covarianța aceloraşi date este

$$\sum_i (x_{1i} - m_1^*)(x_{2i} - m_2^*) = \bar{\rho} \cdot \bar{\sigma}_1 \cdot \bar{\sigma}_2$$

unde $\bar{\rho}$ este corelația.

Ultimele 3 ecuații (4_7) se pot scrie :

$$\begin{aligned} a_{11} &= |a_{kh}| \bar{\sigma}_2^2 \\ a_{22} &= |a_{kh}| \bar{\sigma}_1^2 \\ a_{12} &= -|a_{kh}| \bar{\rho} \cdot \bar{\sigma}_1 \bar{\sigma}_2 \end{aligned} \qquad (4_{10})$$

de aici deducem combinația :

$$a_{11} a_{22} - a_{12}^2 = |a_{kh}|^2 \sigma_1 \cdot \sigma_2 - |a_{kh}|^2 \bar{\rho}^2 \cdot \bar{\sigma}_1^2 \cdot \bar{\sigma}_2^2 = |a_{kh}|^2 \bar{\sigma}_1^2 \cdot \bar{\sigma}_2^2 (1 - \rho^2)$$

sau încă ținînd seama că $a_{11} a_{22} - a_{12}^2 = |a_{kh}|$, găsim relația

$$|a_{kh}| = \frac{1}{\bar{\sigma}_1^2 \cdot \bar{\sigma}_2^2 (1 - \bar{\rho}^2)} \qquad (4_{11}$$

Înlocuind în relația (2_{10}) avem estimațiile :

$$\begin{aligned} a_{11}^* &= \frac{1}{\bar{\sigma}_1^2 (1 - \bar{\rho}^2)} \\ a_{12}^* &= \frac{\bar{\rho}}{\bar{\sigma}_1 \cdot \bar{\sigma}_2 (1 - \rho^2)} \\ a_{22}^* &= \frac{1}{\bar{\sigma}_2^2 (1 - \bar{\rho}^2)} \end{aligned} \qquad (4_{12})$$

Înlocuind aceste estimații împreună cu (4_8) în (4_2), se obține densitatea populației Γ cu ajutorul datelor de selecție.

Aceste rezultate se extind fără dificultate la o populație normală s-dimensională.

§ 6. Distribuția indicatorilor de probabilitate maximă

Să rezolvăm mai întîi o problemă simplă : Se dă o populație Γ normal distribuită cu media m cunoscută și ne propunem să determinăm distribuția indicatorului de probabilitate maximă pentru dispersia distribuției. Vom scrie distribuția sub forma :

$$f(x;\ m,\ \lambda^2) = \frac{1}{\sqrt{2\pi}\sqrt{\lambda^2}} e^{-\frac{1}{2}\frac{(x-m)^2}{\lambda^2}}. \qquad (4_{13})$$

Scanned with OKEN Scanner

Dispersia distribuției este deci λ^2 și trebuie să determinăm distribuția indicatorului său de probabilitate maximă λ^{*2}.

Din populația Γ să extragem o selecție de mărime n care dă valorile $x_1, x_2 \ldots x_n$. Probabilitatea acestei selecții este

$$P = \prod_{i=1}^{n} f(x_i;\ m,\ \lambda^2) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \frac{1}{\lambda^n}\, e^{-\frac{1}{2\lambda^2}\sum_{i=1}^{n}(x_i-m)^2} \qquad (4_{14})$$

deci logaritmul natural al acestei probabilități:

$$\ln P = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\lambda^2 - \frac{1}{2\lambda^2}\sum_{i=1}^{n}{}'(x_i - m)^2 \qquad (4_{15})$$

De aici deducem derivata în raport cu λ^2

$$\frac{\partial \ln P}{\partial \lambda^2} = \frac{1}{2\lambda^2}\left[\frac{1}{\lambda^2}\sum_{i=1}^{n}(x_i - m)^2 - n\right] = 0$$

de unde rezultă valoarea indicatorului

$$\lambda^{*2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2 \qquad (4_{16})$$

Introducînd notația

$$y = \sum_{i=1}^{n}{}'\left(\frac{x_i - m}{\lambda}\right)^2 \qquad (4_{17})$$

relația (4_{16}) se poate scrie amplificînd prin λ^2:

$$\lambda^{*2} = \frac{\lambda^2}{n}\, y \qquad (4'_{16})$$

Se știe însă că variabila y, suma patratelor a n variabile aleatorii independente distribuite normal,

$$\frac{x_i - m}{\lambda}, \qquad 1 \leqslant i \leqslant n \qquad (4_{18})$$

cu media 0 și dispersia 1 are o distribuție χ^2 cu n grade de libertate. Densitatea acestei variabile o cunoaștem; ea este dată de funcția:

$$f(y) = \frac{1}{\Gamma\left(\frac{n}{2}\right)\cdot 2^{\frac{n}{2}}}\cdot y^{\frac{n}{2}-1}\, e^{-\frac{1}{2}y}, \quad y > 0 \qquad (4_{19})$$

Distribuția indicatorului $\lambda^{*2} = \frac{\lambda^2 y}{n} = \lambda^{*2}(y)$ va fi deci:

$$\varphi(\lambda^{*2}) = f\,[y(\lambda^{*2})] \left| \frac{dy}{d\lambda^{*2}} \right|$$

și deoarece $\left| \frac{dy}{d\lambda^{*2}} \right| = \frac{n}{\lambda^2}$ obținem distribuția:

$$\varphi(\lambda^{*2}) = \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}} \frac{n\lambda^{*2}}{\lambda^2}^{\frac{n}{2}-1} e^{-\frac{1}{2}\frac{n\lambda^{*2}}{\lambda^2}} \frac{n}{\lambda^2} =$$

$$= \frac{1}{\Gamma\left(\frac{n}{2}\right)} \left(\frac{n}{2\lambda^2}\right)^{\frac{n}{2}} (\lambda^{*2})^{\frac{n}{2}-1} e^{-\frac{n\lambda^{*2}}{2\lambda^2}} \qquad (4_{20})$$

adică o distribuție Gama în care:

$$a = \frac{n}{2} - 1, \quad b = \frac{2\lambda^2}{n}$$

În mod obișnuit, nu se cunoaște media m a populației normale Γ așa încît, în locul indicatorului $\lambda^{*2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2$ este mai util să se calculeze distribuția indicatorului

$$\lambda_1^{*2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

unde $\bar{x}$ reprezintă media de selecție. Un calcul asemănător cu cel precedent ne conduce la o densitate $\varphi(\lambda_1^{*2})$ analogă cu relația $(2'_{20})$.

Menționăm, fără demonstrație, o proprietate importantă care aparține numai distribuției normale și anume: *media de selecție și dispersia selecțiilor dintr-o populație normală au distribuții independente.* Pentru orice altă distribuție, distribuția dispersiei de selecție nu este independentă de distribuția valorii medii a selecției.

Să considerăm acum o distribuție cu densitatea $f(x; \lambda)$ depinzînd de un singur parametru λ. Indicatorul de probabilitate maximă care estimează pe λ este evident o funcție de valo-

rile $x_1, x_2, \ldots x_n$ ale selecției, fie $\lambda^* (x_1, x_2, \ldots, x_n)$. Vom demonstra că pentru selecții mari, indicatorul $\lambda^* (x_1, x_2, \ldots, x_n)$ are o distribuție normală, în condiții destul de generale.

În acest scop să considerăm funcția aleatorie

$$u(x) = \frac{\partial}{\partial \lambda} \ln f(x, \lambda) \tag{5}$$

sau :

$$u(x) = \frac{1}{f(x, \lambda)} \frac{\partial}{\partial \lambda} f(x, \lambda) \tag{5'}$$

Valoarea medie a funcției u (x) este :

$$M(u) = \int_{-\infty}^{+\infty} u\, f(x, \lambda)\, dx = \int_{-\infty}^{+\infty} \frac{1}{f(x, \lambda)} \left(\frac{\partial}{\partial \lambda} f(x, \lambda) \right) f(x, \lambda)\, dx$$

$$M(u) = \int_{-\infty}^{+\infty} \frac{\partial}{\partial \lambda} f(x, \lambda)\, dx \tag{5_1}$$

Prima ipoteză : Să admitem că operațiile de derivare și integrare în (5_1) sînt permutabile, atunci :

$$M(u) = \frac{\partial}{\partial \lambda} \int_{-\infty}^{+\infty} f(x, \lambda)\, dx$$

și cum $\int_{-\infty}^{+\infty} f(x ; \lambda)\, dx = 1$ căci $f(x ; \lambda)$ este o densitate de probabilitate, rezultă :

$$M(u) = \frac{\partial}{\partial \lambda} (1) = 0 \tag{$5_1'$}$$

Să calculăm acum dispersia funcției u (x). Vom avea :

$$\sigma_u^2 = \int_{-\infty}^{+\infty} [u - M(u)]^2 f(x, \lambda)\, dx = \int_{-\infty}^{+\infty} \left[\frac{1}{f(x, \lambda)} \cdot \frac{\partial}{\partial \lambda} f(x, \lambda) - 0 \right]^2 f(x, \lambda)\, dx$$

sau încă :

$$\sigma_u^2 = \int_{-\infty}^{+\infty} \left[\frac{\partial}{\partial \lambda} f(x\, \lambda) \right]^2 \frac{1}{f(x, \lambda)}\, dx. \tag{5_2}$$

Scriind valoarea medie M (u) sub forma :

$$M(u) = \int_{-\infty}^{+\infty} u(x) f(x, \lambda)\, dx = \int_{-\infty}^{+\infty} \left[\frac{\partial}{\partial \lambda} \ln f(x, \lambda) \right] f(x, \lambda)\, dx$$

Scanned with OKEN Scanner

prin derivare, în raport cu λ, obținem, ținînd seama că M (u) este o constantă :

$$\frac{\partial}{\partial\lambda}\int_{-\infty}^{+\infty}\left[\frac{\partial}{\partial\lambda}\ln f(x;\ \lambda)\right] f(x,\ \lambda)\,dx = \int_{-\infty}^{+\infty}\left[\frac{\partial^2}{\partial\lambda^2}\ln f(x,\lambda)\right] f(x,\ \lambda)\ dx +$$

$$+\int_{-\infty}^{+\infty}\left[\frac{\partial}{\partial\lambda}\ln f(x,\ \lambda)\right]\frac{\partial f(x,\ \lambda)}{\partial\lambda}\ dx = 0$$

sau încă :

$$\int_{-\infty}^{+\infty}\left[\frac{\partial^2}{\partial\lambda^2}\ln f(x,\lambda)\right] f(x,\ \lambda)\ dx + \int_{-\infty}^{+\infty}\frac{1}{f(x,\lambda)}\left[\frac{\partial f(x,\ \lambda)}{\partial\lambda}\right]^2 dx = 0$$

Ultima integrală din prima parte este tocmai σ_u^2 după relația (5_2), deci putem scrie σ_u^2 sub forma :

$$\sigma_u^2 = -M\left(\frac{\partial^2}{\partial\lambda^2}\ln f(x,\ \lambda)\right) \qquad (5_3)$$

Din populația Γ să extragem o selecție de mărime n care dă valorile $x_1, x_2, \ldots x_n$. Valorile funcției u sînt respectiv :

$$u_k = u\ (x_k) = \frac{\partial}{\partial\lambda}\ln f(x_k, \cdot), \quad k = 1,\ 2, \ldots, n \qquad (5_4)$$

Media acestor valori

$$\bar{u} = \frac{1}{n}\sum_{k=1}^{n} u_k = \frac{1}{n}\sum_{k=1}^{n}\frac{\partial}{\partial\lambda}\ln f(x_k,\ \lambda) \qquad (5_5)$$

după cum se știe, are o distribuție normală pentru selecții mari. Valoarea medie a distribuției este media populației Γ, adică $M(\bar{u}) = 0$ și dispersia $\sigma_{\bar{u}}^2 = \frac{\sigma_u^2}{n}$.

Probabilitatea ca să obținem selecția $x_1, x_2, \ldots, x_n$ este egală cu produsul

$$f(x_1,\ \lambda)\cdot f(x_2,\ \lambda)\ldots f(x_n,\ \lambda) \qquad (5_6)$$

de unde deducem logaritmul natural pe care-l notăm prin L :

$$L = \ln f(x_1,\ \lambda) + \ln f(x_2,\ \lambda) + \ldots + \ln f(x_n,\ \lambda) \qquad (5_7)$$

Scanned with OKEN Scanner

Derivînd în raport cu λ obținem :

$$\frac{\partial L}{\partial \lambda} = \frac{\partial}{\partial \lambda} \ln f(x_1, \lambda) + \frac{\partial}{\partial \lambda} \ln f(x_2, \lambda) + \ldots + \frac{\partial}{\partial \lambda} \ln f(x_n, \lambda) \qquad (5_8)$$

Formulele (5_4) și (5_5) ne conduc la relația

$$\bar{u} = \frac{1}{n} \frac{\partial L}{\partial \lambda} \qquad (5_9)$$

Deoarece $\frac{\partial L}{\partial \lambda} = n\,\bar{u}$, rezultă că $\frac{\partial L}{\partial \lambda}$ are o distribuție normală cu media zero și dispersia $D^2(n\,\bar{u}) = n^2 D^2(\bar{u}) = n\,\sigma_u^2$.

Să găsim acum distribuția indicatorului de probabilitate maximă λ^*. În acest scop vom considera următoarele ipoteze :

1°. λ este o rădăcină a ecuației $\frac{\partial L(\lambda)}{\partial \lambda} = 0$.

2°. Derivata $\frac{\partial L(\lambda^*)}{\partial \lambda}$ se poate dezvolta în serie Taylor după puterile lui $(\lambda^* - \lambda)$ adică :

$$\frac{\partial L(\lambda^*)}{\partial \lambda} = \frac{\partial L(\lambda)}{\partial \lambda} + \frac{\partial^2 L(\lambda)}{\partial \lambda^2}(\lambda^* - \lambda) + \frac{1}{2}\frac{\partial^3 L(\lambda^*)}{\partial \lambda^3}(\lambda^* - \lambda)^2 \qquad (5_{10})$$

unde $\lambda' \in (\lambda, \lambda^*)$.

În baza ipotezei 1° putem scrie :

$$\frac{\partial L(\lambda)}{\partial \lambda} = -\frac{\partial^2 L(\lambda)}{\partial \lambda^2}(\lambda^* - \lambda) - \frac{1}{2}\frac{\partial^2 L(\lambda')}{\partial \lambda^3}(\lambda^* - \lambda)^2 \qquad (5_{11})$$

Cum variabila aleatorie $\frac{\partial L(\lambda)}{\partial \lambda}$ are o distribuție normală pentru n mare, și variabila normată $\frac{1}{\sqrt{n}\,\sigma_u} \cdot \frac{\partial L(\lambda)}{\partial \lambda}$ are o distribuție normală cu valoarea medie zero și dispersia 1.

Din relația (5_{11}) deducem înmulțind prin $1/\sigma_u \sqrt{n}$:

$$\frac{1}{\sigma_u \sqrt{n}} \frac{\partial L(\lambda)}{\partial \lambda} = -\frac{1}{\sigma_u \sqrt{n}} \cdot \frac{\partial^2 L(\lambda)}{\partial \lambda^2}(\lambda^* - \lambda) - \frac{1}{\sigma_u \sqrt{n}} \cdot \frac{1}{2}$$

$$\frac{\partial^3 L(\lambda')}{\partial \lambda^3}(\lambda^* - \lambda)^2.$$

Scanned with OKEN Scanner

Făcînd substituția $z = \sqrt{n} \cdot \sigma_u (\lambda^* - \lambda)$ (5_{12})

obținem :

$$\frac{1}{\sigma_u \sqrt{n}} \frac{\partial L(\lambda)}{\partial \lambda} = - z \frac{1}{\sigma_u^2} \left[\frac{1}{n} \frac{\partial^2 L(\lambda)}{\partial \lambda^2} \right] - \frac{z^2}{\sigma_u^3 \sqrt{n}} \left[\frac{1}{n} \cdot \frac{1}{2} \frac{\partial^3 L(\lambda')}{\partial \lambda^3} \right] \quad (5_{13})$$

Din relația (5_8) rezultă :

$$\frac{\partial^2 L(\lambda)}{\partial \lambda^2} = \sum_{k=1}^{n} \frac{\partial^2}{\partial \lambda^2} \ln f(x_k, \lambda)$$

deci :

$$\frac{1}{n} \frac{\partial^2 L(\lambda)}{\partial \lambda^2} = \frac{1}{n} \sum_{k=1}^{n} \frac{\partial^2}{\partial \lambda^2} \ln f(x_k, \lambda)$$

așa încît prima paranteză din partea a doua reprezintă media de selecție a funcției $\frac{\partial^2}{\partial \lambda^2} f(x, \lambda)$. Dar valoarea acestei funcții este egală cu $-\sigma_u^2$ (relația 5_3). Dacă dispersia $D^2 \left[\frac{\partial^2}{\partial \lambda^2} \ln f(x, \lambda) \right.$ este finită, în baza inegalității lui Cebîșev deducem că $\frac{1}{n} \frac{\partial^2 L(\lambda)}{\partial \lambda^2}$ tinde către $-\sigma_u^2$ cu o probabilitate tinzînd către 1 cînd $n \to \infty$. În consecință pentru n mare, primul termen din partea a doua este aproximativ egal cu z. Al doilea termen din partea a doua tinde către zero pentru n mare dacă $\frac{\partial^3 L(\lambda')}{\partial \lambda^3}$ este finită. Cum partea întîi are o distribuție asimptotic normală cu media zero și dispersia 1 și partea a doua, adică z va avea aceeași distribuție. Atunci variabila λ^* dată de relația (5_{12}) :

$$\lambda^* = \lambda + \frac{z}{\sigma_u \sqrt{n}} \quad (5_{14})$$

va avea pentru selecții mari o distribuție asimptotic normală cu valoarea medie :

$$M(\lambda^*) = M(\lambda) + \frac{1}{\sigma_u \sqrt{n}} M(z) = \lambda$$

și dispersia :

$$D^2(\lambda^*) = \frac{1}{n \sigma_u^2} \quad (5_{15})$$

(λ reprezintă adevărata valoare a parametrului).

Scanned with OKEN Scanner

Din relațiile (5_{14}) și (5_{15}) putem scrie :

$$M(\lambda^*) = \lambda + \varphi(n)$$

$$D^2(\lambda^*) \underset{n\to\infty}{\longrightarrow} 0 \quad \text{și} \quad \varphi(n) \underset{n\to\infty}{\longrightarrow} 0$$

Prin urmare, indicatorul de probabilitate maximă λ^* estimează în general numai corect pe λ. În consecință, λ reprezintă o medie de selecții mari, adică valoarea către care tinde media de selecție cînd $n \to \infty$.

Pentru o clasă întinsă de distribuții, indicatorul de probabilitate maximă are o distribuție aproximativ normală în jurul valorii exacte a parametrului, ca medie de selecții mari.

Rezultatele stabilite mai sus sînt aplicabile atît distribuțiilor discrete, cît și celor continue, singura diferență în raționamente fiind înlocuirea integralelor prin sume.

Concluzii. Indicatorul de probabilitate maximă λ^* (x_1; x_2, ... $\lambda^* x_n$) are o distribuție normală pentru selecții mari, dacă densitatea $f(x, \lambda)$ a populației Γ satisface următoarele condiții :

1°. Este posibilă permutabilitatea operațiilor de integrare în raport cu x și derivare în raport cu λ ;

2°. $M(u) = M\left[\frac{\partial}{\partial\lambda} \ln f(x, \lambda)\right] = 0$;

3°. Dacă $M \frac{\partial^2}{\partial\lambda^2} \ln f(x, \lambda)$ și $D^2\left[\frac{\partial^2}{\partial\lambda^2} \ln f(x, \lambda)\right]$

sînt finite ;

4°. Dacă $\frac{1}{n} \frac{\partial^3 L(\lambda)}{\partial\lambda^3}$ este mărginită pentru orice λ.

5°. $\frac{\partial L(\lambda)}{\partial\lambda} = 0$ pentru valoarea sa maximă.

Raționamentele din cazul distribuțiilor cu un singur parametru se extind și la distribuțiile depinzînd de mai mulți parametri. Astfel, pentru distribuția $f(x, \lambda_1, \lambda_2, \ldots \lambda_r)$, depinzînd de r parametri, indicatorii de probabilitate maximă $\lambda_1^*, \lambda_2^*, \ldots \lambda_r^*$, pentru selecții mari, au aproximativ o distribuție normală r-dimensională cu valorile medii $\lambda_1, \lambda_2, \ldots \lambda_r$ și coeficienții formei pătratice de la exponent egali cu elementele matricei $\| n\, a_{kh} \|$ unde :

$$a_{kh} = -M\left[\frac{\partial^2}{\partial\lambda_k\, \partial\lambda_h} \ln f(x; \lambda_1 \lambda_2, \ldots, \lambda_r)\right] \qquad (5_{16})$$

Aplicație. Să găsim distribuția indicatorilor de probabilitate maximă ai celor doi parametri din distribuția normală scrisă sub forma următoare :

$$f(x, \lambda_1, \lambda_2) = \frac{1}{\sqrt{2\pi\lambda_2}} e^{-\frac{(x-\lambda_1)^2}{2\lambda_2}} \tag{6}$$

Dacă n este mărimea selecției, știm că indicatorii λ_1 și λ_2 sînt dați de relațiile :

$$\lambda_1^* = \frac{1}{n} \sum_{k=1}^{n} x_k$$
$$\lambda_2^* = \frac{1}{n} \sum_{k=1}^{n} (x_k - \overset{*}{\lambda_1})^2 \tag{6_1}$$

După teorema precedentă λ_1^* și λ_2^* au distribuții normale cu mediile λ_1 și λ_2 și coeficienții n a_{kh} în forma pătratică, unde

$$a_{kh} = -M\left[\frac{\partial^2}{\partial\lambda_1 \partial\lambda_2} \ln f\right] \tag{6_2}$$

Cum

$$\ln f = -\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \lambda_2 - \frac{1}{2\lambda_2} (x - \lambda_1)^2$$

vom avea derivatele :

$$\frac{\partial^2}{\partial\lambda_1^2} \ln f = -\frac{\lambda_1}{\lambda_2}$$
$$\frac{\partial^2}{\partial\lambda_1 \partial\lambda_2} \ln f = -\frac{x - \lambda_1}{\lambda_2^2} \tag{6_3}$$
$$\frac{\partial^2}{\partial\lambda_2^2} \ln f = \frac{1}{2\lambda_2^2} - \frac{(x - \lambda_1)^2}{\lambda_2^3}$$

Deoarece $M(x) = \lambda_1$, $M(x - \lambda_1)^2 = \lambda_2$, rezultă :

$$\| a_{kh} \| = \begin{Vmatrix} \frac{\lambda_1}{\lambda_2} & 0 \\ 0 & \frac{1}{2\lambda_2^2} \end{Vmatrix}$$

Scanned with OKEN Scanner

Prin urmare, distribuția de selecție mare a indicatorilor λ^*_1 și λ^*_2 este :

$$h\left(\lambda_1^*, \lambda_2^*\right) = \frac{1}{2\pi} \cdot \frac{n}{\sqrt{2\lambda_2^3}} e^{-\frac{n}{2}\left[\frac{\left(\lambda_1^*-\lambda_1\right)^2}{\lambda_2} + \frac{\left(\lambda_2^*-\lambda_2\right)}{2\lambda_2^2}\right]} \tag{6}$$

Distribuția lui λ_1^* pentru selecții mari este cea normală, iar distribuția lui λ_2^* pentru selecții de orice mărime este o distribuție Gama.

§ 7. Estimații pentru selecțiile nerepetate

Ne-am ocupat pînă acum de estimațiile făcute pe baza unor selecții extrase din populațiile infinite în care caz problema selecției repetate sau nerepetate nu se mai pune, precum și de selecțiile repetate dintr-o populație finită.

Pentru aceste selecții am stabilit formulele : $M(\bar{x}) = m$, $D(\bar{x}) = \frac{\sigma^2}{n}$ unde $\bar{x}$ este ca de obicei media de selecție, iar m și σ^2 reprezintă valoarea probabilă și respectiv dispersia variabilei aleatorii x definită pe populația Γ.

Pentru valori mari ale volumului n al selecției, variabila $\bar{x}$ are o distribuție normală, iar pentru selecțiile mici ea are o distribuție normală numai în cazul cînd variabila x are și ea o distribuție normală.

În acest paragraf ne vom ocupa de unele probleme de estimație pentru selecțiile nerepetate extrase dintr-o populație finită. Să considerăm populația Γ compusă din A elemente pe care le presupunem distincte între ele ; pentru fiecare element ne interesează o anumită caracteristică. Așa, de pildă, dacă populația Γ este formată din 90 bile marcate de la 1 la 90 și considerăm numărul scris pe fiecare bilă ca o variabilă, aceasta poate lua valorile $x = 1, 2, 3 \ldots, 89, 90$. În general, să notăm prin :

$$x_1, x_2, \ldots x_A \tag{7}$$

valorile caracteristicii x care ne interesează, corespunzătoare celor A elemente ale populației Γ.

Scanned with OKEN Scanner

Valoarea probabilă și dispersia variabilei x vor fi :

$$m = \frac{1}{A} \sum_{k=1}^{A} x_k \tag{7_1}$$

$$\sigma^2 = \mu_2 = \frac{1}{A} \sum_{k=1}^{A} (x_k - m)^2 \tag{7_2}$$

Să efectuăm acum o selecție de mărime n, nerepetată și pe baza acestei selecții să estimăm media m a populației Γ. Admițînd că toate elementele din Γ au aceeași probabilitate p de a face parte din selecție, după schema bilei nerevenite în care facem a = 1, x = 1, m = A (capitolul III), găsim :

$$p = \frac{C_a^x \, C_{m-a}^{n-x}}{C_m^n} = \frac{C_1^1 \, C_{A-1}^{n-1}}{C_A^n} = \frac{n}{A} \tag{7_3}$$

În mod analog, făcînd a = 2, x = 2 se găsește probabilitatea ca două elemente oarecare să facă parte din selecție ; notînd cu $p^{(2)}$ această probabilitate avem ;

$$p^{(2)} = \frac{C_{A-2}^{n-2}}{C_A^n} = \frac{n\,(n-1)}{A\,(A-1)} \tag{7_3'}$$

Să atașăm fiecărui element x_k, o variabilă aleatorie t_k, $(1 \leqslant k \leqslant A)$ cu valorile 1 sau 0 după cum elementul t_k a fost sau nu extras în selecție. Putem deci scrie relațiile :

$$P(\bar{t}_k = 1) = \frac{n}{A}, \quad P(t_k = 0) = 1 - \frac{n}{A}. \tag{7_4}$$

Variabilele aleatorii $t_1, t_2, \ldots t_A$ satisfac relația evidentă

$$t_1 + t_2 + \ldots + \bar{t}_A = n \tag{7_5}$$

ceea ce înseamnează că n din ele iau valoarea 1, iar (A — n) iau valoarea O. Media de selecție $\bar{x}$ se prezintă ca o medie a valorilor x_k cu ponderile t_k, deci :

$$\bar{x} = \frac{\sum_{k=1}^{A} t_k x_k}{\sum_{k=1}^{A} t_k} = \frac{1}{n} \sum_{k=1}^{A} t_k x_k \tag{7_6}$$

Să estimăm media m cu ajutorul mediei de selecție $\bar{x}$. Pentru aceasta vom calcula valoarea probabilă M ($\bar{x}$) și dispersia D^2 ($\bar{x}$) a mediei de selecție, considerată după relația (7_6) ca o variabilă aleatorie. Mai întîi, însă, vom observa că pentru orice k = 1, 2, ... A avem :

$$M(\bar{t}_k) = 1 \cdot p + 0\,(1 - p) = p \qquad (7_7)$$

$$M(t_k^2) = 1^2\, p + 0^2 \cdot (1 - p) = p \qquad (7'_7)$$

iar pentru orice k și h :

$$M(t_k\, t_h) = \frac{n\,(n-1)}{A\,(A-1)}. \qquad (7''_7)$$

În adevăr, variabilele t_k și t_h, în baza relației (7_5), sînt dependente ; produsul $t_k\, t_h$ poate lua valorile :

$$1 \cdot 1 = 1;\quad 1 \cdot 0 = 0;\quad 0 \cdot 1 = 0;\quad 0 \cdot 0 = 0$$

cu probabilitățile respective p_{11}, p_{10}, p_{01} și p_{00}.
Cum $p_{11} = p^{(2)}$, adică probabilitatea ca perechea x_k și x_h să ajungă în selecție, rezultă $M(t_k\, t_h) = 1 \cdot 1\, p_{11} = p^{(2)} = \frac{n\,(n-1)}{A\,(A-1)}$, după relația ($7'_3$).

Cu aceste pregătiri, valoarea probabilă M ($\bar{x}$) se scrie :

$$M(\bar{x}) = \frac{1}{n} \sum_{k=1}^{A} x_k\, M(t_k) = \frac{1}{n} \sum_{k=1}^{A} x_k \cdot p = \frac{p}{n} \sum_{k=1}^{A} x_k$$

sau încă, înlocuind pe p cu raportul n/A găsim :

$$M(\bar{x}) = \frac{1}{m} \sum_{k=1}^{m} x_k = m \qquad (7_8)$$

Dispersia variabilei $\bar{x}$ se calculează cu formula cunoscută :

$$D^2(\bar{x}) = \sigma_{\bar{x}}^2 = M(\bar{x} - m)^2 = M\left[\frac{\sum_1^A x_k\, t_k}{\sum_1^A t_k} - m\right]^2 = M\left[\frac{\sum_1^A t_k\,(x_k - m)}{n}\right]^2 =$$

$$= \frac{1}{n^2} M\left[\sum_1^A (x_k - m)^2\, t_k^2 + \sum_{k \neq h} (x_k - m)(x_h - m)\, t_k\, t_h\right] =$$

$$= \frac{1}{n^2} \sum_1^A (x_k - m)^2\, M(t_k^2) + \frac{1}{n^2} \sum_{k \neq h} (x_k - m)(x_h - m)\, M(t_k\, t_h)$$

Scanned with OKEN Scanner

Ținînd seama de relațiile $(7'_7)$ și $(7''_7)$ găsim :

$$D^2(\bar{x}) = \frac{1}{n\,A}\sum_{1}^{A}(x_k - m)^2 + \frac{(n-1)}{n\,A\,(A-1)}\sum_{k\neq h}(x_k - m)(x_h - m)$$

Primul termen din partea dreaptă este $\frac{1}{n}\sigma^2$, în baza relației (7_2). În ce privește al doilea vom scrie succesiv :

$$\sum_{k\neq h}(x_k - m)(x_h - m) = \sum_{k}{}'\left[(x_k - m)\sum_{h\neq k}(x_h - m)\right] =$$

$$= \sum_{k}(x_k - m)\left[(x_1 - m) + \ldots + (x_{k-1} - m) + (x_{k+1} - m) + \ldots + (x_A - m)\right]$$

Adunînd în paranteza mare și scăzînd $(x_k - m)$ obținem :

$$\sum_{k}(x_k - m)\left[\sum_{h=1}^{A}(x_h - m) - (x_k - m)\right] = \sum_{k}(x_k - m)\left[\sum_{h=1}^{A}x_h - mA - (x_k - m)\right]$$

Dar $\sum_{h=1}^{A} x_k = m\,A$ după relația (7_1), deci :

$$\sum_{k\neq h}(x_k - m)(x_h - m) = -\sum_{k=1}^{A}(x_k - m)^2 = -A\,\sigma^2$$

Așadar, vom avea dispersia :

$$D^2(\bar{x}) = \sigma^2_{\bar{x}} = \frac{\sigma^2}{n} - \frac{n-1}{n\,(A-1)}\sigma^2 = \frac{A-n}{A-1}\frac{\sigma^2}{n} \qquad (7_3)$$

Abaterea medie pătratică va fi :

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{1-\frac{n}{A}}{1-\frac{1}{A}}}. \qquad (7'_3)$$

Deoarece factorul

$$\sqrt{\left(1-\frac{n}{A}\right):\left(1-\frac{1}{A}\right)} < 1$$

rezultă $\sigma_{\bar{x}} < \frac{\sigma}{\sqrt{n}}$ pe care am găsit-o în cazul selecției repetate.

Exemplu. Dintr-un lot de 1 600 piese se face o selecție de 144 piese pentru care se stabilește media unei anumite dimensiuni

Scanned with OKEN Scanner

de 15,42 cm și abaterea medie pătratică de 0,16 mm. Să se compare abaterea medie pătratică $\sigma_{\bar{x}}$ pentru selecțiile repetate și nerepetate.

Cu datele problemei avem :

$$\bar{x} = 15{,}42\,;\quad \sigma \approx 0{,}16\,;\quad n = 144\,:\quad A = 1\,600.$$

Pentru selecția repetată avem :

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0{,}16}{\sqrt{144}} = 0{,}013 \text{ mm}$$

iar pentru selecția nerepetată

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{1 - \frac{n}{A}}{1 - \frac{1}{A}}} = \frac{0{,}16}{\sqrt{144}} \times 0{,}96 = 0{,}0128 \text{ mm}$$

Observație. În anumite împrejurări, caracteristica x nu mai este cantitativă, ci calitativă. De pildă, populația Γ este formată din A bile din care a sînt albe și restul (A—a) sînt negre. Fiecărei bile îi vom atașa o variabilă t_k ce poate lua valorile 1 sau 0 după cum bila este albă sau nu. Notînd bilele albe de la 1 la a, iar cele negre de la (a + 1) la A avem :

$$t_1 = t_2 = \ldots = t_a = 1\,;\ t_{a+1} = t_{a+2} = \ldots = t_A = 0.$$

De unde deducem :

$$m = \frac{1}{A} \sum_{k=1}^{A} t_k = \frac{1}{A} \left[1 \cdot a + 0\,(A - a)\right] = \frac{a}{A} \qquad (8)$$

adică media este egală cu frecvența relativă a numărului de bile albe din urnă. Dispersia numărului acestor bile este :

$$\sigma^2 = \frac{a}{A}\left(1 - \frac{a}{A}\right)^2 + \frac{A - a}{A}\left(0 - \frac{a}{A}\right)^2 = \frac{a}{A}\left(1 - \frac{a}{A}\right)$$

Notînd a/A = p și înlocuind în ($7'_9$) găsim :

$$\sigma_{\frac{a}{A}} = \sqrt{\frac{p\,(1 - p)}{n}} \sqrt{\frac{1 - \frac{n}{A}}{1 - \frac{1}{A}}} \qquad (8_1)$$

Scanned with OKEN Scanner

Aşadar, siguranţa estimării probabilităţii cu frecvenţa relativă dată de selecţie depinde de mărimea $\sigma_{\frac{a}{A}}$.

Exemplu. Dintr-un lot de 2 500 chei se face o selecţie de 225; din acestea 10 sînt defecte. Să se calculeze abaterea tip pentru frecvenţa relativă.

Luînd pentru populaţia de 2 500 chei frecvenţa din selecţie de $10/225 = 4\%$ găsim:

$$\sigma_{\frac{a}{A}} = \sqrt{\frac{0{,}04\ (1 - 0{,}04)}{225}} \sqrt{\frac{1 - \frac{225}{2\,500}}{1 - \frac{1}{2\,500}}} = 1\,\%$$

Dacă nu se ţine seama de volumul $A = 2\,500$ al populaţiei se găseşte:

$$\sigma_{\frac{a}{A}} = \sqrt{\frac{0{,}04\ (1 - 0{,}04)}{225}} \approx 1{,}3\,\%$$

Capitolul XI

ESTIMAREA INTERVALELOR

§ 1. Intervale de confidență

În capitolul precedent ne-am ocupat de unele probleme în legătură cu estimația parametrilor unei distribuții statistice. Deoarece estimația este o evaluare apropiată a adevăratei valori a parametrului, se pune în mod natural problema calculului erorii probabile care afectează estimația făcută. Cu alte cuvinte, dacă am estimat parametrul λ găsind estimația λ^*, să determinăm și un interval $(\lambda^* - h, \lambda^* + h)$ în jurul punctului λ^* așa fel ca parametrul adevărat $\lambda \in (\lambda^* - h, \lambda^* + h)$, precum și o măsură de încredere că λ este conținut în acest interval.

De exemplu o fabrică de textile, ținînd seama de costurile de producție în sine, de proporția aferentă a salariilor, a amortizărilor etc..., estimează costul unui metru de stofă la $225 \pm 13{,}70$ lei; aceasta înseamnă că prețul real al unui metru de stofă este cu foarte mare probabilitate între 211,30 și 238,70 lei.

§ 2. Prima metodă (elementară) pentru determinarea intervalelor de confidență

Această metodă constă în *determinarea unei funcții de datele selecției și parametrul de estimat și care are o distribuție independentă de parametru.*

Un exemplu simplu va lămuri modul de aplicare a acestei metode: Se consideră o populație normală Γ cu media m necunoscută și dispersia $\sigma^2 = 16$. Din această populație se extrage o selecție de trei elemente și se obțin valorile:

$$x_1 = 1{,}2\,;\ x_2 = 1{,}4\,;\ x_3 = 1{,}3$$

Se știe că indicatorul de probabilitate maximă pentru media necunoscută m a colectivității Γ este media de selecție $m = \bar{x} = 1{,}3$. Să găsim acum o margine inferioară și o margine superioară între care este cuprinsă aproape sigur valoarea m a mediei adevărate. În acest scop, folosim abaterea normată:

$$\bar{\eta} = \frac{\bar{x} - m}{\left(\frac{\sigma}{\sqrt{n}}\right)} \tag{1}$$

care, după cum s-a văzut, are o distribuție normală cu media $M(\bar{\eta}) = 0$ și dispersia $D^2(\bar{\eta}) = 1$. În exemplul de mai sus avem $n = 3$, $\sigma = \sqrt{16} = 4$, deci $\bar{\eta} = \frac{\bar{x} - m}{\frac{4}{\sqrt{3}}}$. Densitatea lui $\bar{\eta}$ este:

$$f(\bar{\eta}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\bar{\eta}^2} \tag{1_1}$$

și ea nu depinde de parametrul m. Variabila $\bar{\eta}$ este funcția căutată.

Graficul acestei distribuții este următorul:

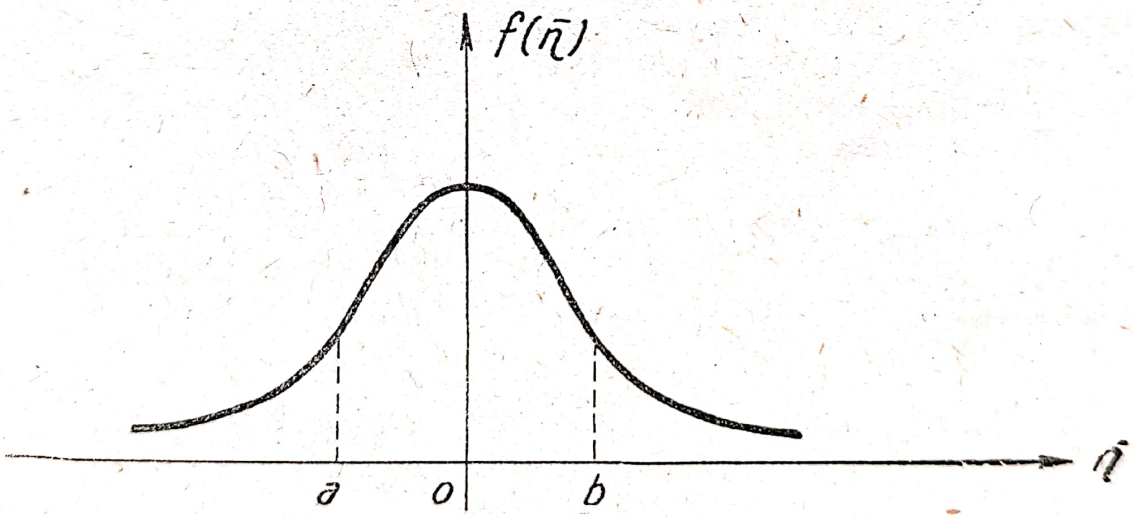

Fig. 49

Aria mărginită de această curbă este:

$$A = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\bar{\eta}^2} d\bar{\eta} = 1 \tag{1_2}$$

deoarece $f(\bar{\eta})$ este o densitate de probabilitate.

Scanned with OKEN Scanner

Probabilitatea ca $\bar{\eta}$ să fie cuprinsă între două numere η' și $\bar{\eta}''$ este egală cu :

$$P(\bar{\eta}' < \bar{\eta} < \bar{\eta}'') = \int_{\eta'}^{\bar{\eta}''} f(\bar{\eta})\ d\bar{\eta} \qquad (1_3)$$

și reprezintă aria mărginită de f $(\bar{\eta})$ și ordonatele punctelor $\bar{\eta}'$ și $\bar{\eta}''$.

Așa de exemplu, pentru $\eta' = -1{,}96$, $\eta'' = 1{,}96$ putem scrie :

$$P(-1{,}96 < \bar{\eta} < 1{,}96) = \int_{-1{,}96}^{1{,}96} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\eta^2}\ d\bar{\eta} = 2\Phi(1{,}96)$$

În tabela I găsim $2\Phi(1{,}96) = 0{,}95$, deci :

$$P\left[-1{,}96 < \frac{\bar{x} - m}{\left(\frac{4}{\sqrt{3}}\right)} < 1{,}96\right] = 0{,}95 \qquad (1_4)$$

De aici rezultă :

$$-1{,}96\,\frac{4}{\sqrt{3}} < \bar{x} - m\,;\quad \bar{x} - m < 1{,}96\,\frac{4}{\sqrt{3}}$$

sau încă :

$$m < \bar{x} + 1{,}96\,\frac{4}{\sqrt{3}}\,;\quad m > \bar{x} - 1{,}96\,\frac{4}{\sqrt{3}}$$

Deci relația (1_4) se poate scrie :

$$P\left(\bar{x} - 1{,}96\,\frac{4}{\sqrt{3}}\,m < \bar{x} + 1{,}96\frac{4}{\sqrt{3}}\right) = 0{,}95 \qquad (1'_4)$$

Media de selecție $\bar{x}$ variază de la o selecție la alta, deci ea este o variabilă aleatorie ; în consecință, și intervalul :

$$\left(\bar{x} = 1{,}96\,\frac{4}{\sqrt{3}},\ \bar{x} + 1{,}96\,\frac{4}{\sqrt{3}}\right) = (\bar{x} - 4{,}53,\ x + 4{,}53)$$

este un interval aleator. În acest interval se va găsi adevărata valoare a parametrului m cu o probabilitate de 0,95. Aceasta înseamnă că dacă se va extrage un mare număr de selecții de cîte trei elemente și se va calcula pentru fiecare selecție

Scanned with OKEN Scanner

intervalul de mai sus, în medie 95% din acestea vor conține media m a populației Γ.

În particular, înlocuind $\bar{x}$ prin valoarea 1,3 dată de selecție se obține intervalul $(-3{,}23\,;\ 5{,}83)$. Acest interval se numește *interval de confidență 95%*. Avem relația: $P(-3{,}23 < m < 5{,}83) = 0{,}95$. Trebuie să observăm că există mai multe intervale de confidență 95%: Orice numere a și b ale căror ordonate închid 95% din aria situată sub curba $f(\bar{\eta})$ din fig. 50 determină un interval de confidență 95%. Noi căutăm însă intervalul de confidență cel mai scurt; intervalul minim (a, b) se obține pentru $f(a) = f(b)$, adică punctele a și b sînt simetrice față de axa $O\bar{\eta}$. În adevăr, avem probabilitatea $P[a < \bar{\eta} < b] = \int_a^b f(\bar{\eta})\, d\bar{\eta} = 0{,}95$ sau încă, scriind sub forma echivalentă:

$$P\left[\bar{x} - b\frac{\sigma}{\sqrt{n}} < m < \bar{x} - a\frac{\sigma}{\sqrt{n}}\right] = 0{,}95$$

Lungimea intervalului este:

$$x - a\frac{\sigma}{\sqrt{n}} - \left(\bar{x} - b\frac{\sigma}{\sqrt{n}}\right) = (b - a)\frac{\sigma}{\sqrt{n}}$$

Minimul acestui interval se obține, după cum se știe, formînd funcția:

$$\varphi(a, b) = \frac{\sigma}{\sqrt{n}}(b - a) + \lambda \int_a^b f(\bar{\eta})\, d\bar{\eta}$$

din care deducem:

$$\varphi'_a = -\frac{\sigma}{\sqrt{n}} - \lambda f(a) = 0$$

$$\varphi'_b = \frac{\sigma}{\sqrt{n}} + \lambda f(b) = 0$$

Eliminînd pe λ din aceste două relații se găsește:

$$\lambda = -\frac{\frac{\sigma}{\sqrt{n}}}{f(a)} = -\frac{\frac{\sigma}{\sqrt{n}}}{f(b)}$$

de unde rezultă $f(a) = f(b)$.

Scanned with OKEN Scanner

Prin urmare, intervalul $(-3{,}23\,;\ 5{,}83)$ este cel mai scurt din toate intervalele de confidență 95%.

Observație. Metoda se aplică la estimarea mai multor parametri, însă, în acest caz se obține o *regiune de confidență.* Așa de exemplu, pentru cei doi parametri m și σ^2 din distribuția normală, o regiune de confidență 97% va fi o regiune obținută din selecție în așa fel că dacă s-ar extrage un număr mare de selecții din populația normală și s-ar construi pentru fiecare selecție o regiune de confidență, în medie 97% din ele vor conține punctul ale cărui coordonate sînt valorile exacte m și σ^2 ale parametrilor.

Aplicația I. Intervale de confidență pentru valoarea medie a unei populații normale. În cazul cînd se cunoaște abaterea medie pătratică σ a populației normale Γ, limitele intervalului de confidență 95%, de pildă, se calculează din relația $P\left(\bar{x}-1{,}96\,\frac{\sigma}{\sqrt{n}}<m<\bar{x}+1{,}96\,\frac{\sigma}{\sqrt{n}}\right)=0{,}95$. Pentru o anumită selecție se cunoaște volumul n și media de selecție $\bar{x}$, așa încît, putem afirma cu o probabilitate de 0,95 că media m a colectivității Γ se găsește în intervalul $\left(\bar{x}-1{,}96\,\frac{\sigma}{\sqrt{n}},\ \bar{x}+1{,}96\,\frac{\sigma}{\sqrt{n}}\right)$ în care toate elementele, inclusiv σ, sînt cunoscute.

De obicei, nu se cunoaște σ, iar limitele intervalului de confidență nu pot fi determinate în modul indicat mai sus. Problema astfel pusă poate fi rezolvată introducînd variabila aleatorie.

$$t=\frac{\bar{x}-m}{\sqrt{\dfrac{\sum_{k=1}^{n}(x_k-\bar{n})^2}{n\,(n-1)}}} \qquad (1_5)$$

pe care o putem scrie sub forma:

$$t=\frac{\dfrac{\bar{x}-m}{\sigma}}{\sqrt{\dfrac{\dfrac{1}{n}\sum_{k=1}^{n}\dfrac{(x_k-\bar{x})^2}{\sigma}}{n-1}}} \qquad (1'_5)$$

Se știe că variabila $\frac{\bar{x} - m}{\sigma}$ are o distribuție normală cu media

$$M\left(\frac{\bar{x} - m}{\sigma}\right) = \frac{1}{\sigma}(M\bar{x} - m) = \frac{1}{\sigma}(m - m) = 0$$

și dispersia :

$$D^2\left(\frac{\bar{x} - m}{\sigma}\right) = \frac{1}{\sigma^2} \quad D^2(\bar{x} - m) = \frac{1}{\sigma^2} D^2(\bar{x}) = \frac{1}{\sigma^2} \cdot \frac{\sigma^2}{n} = \frac{1}{n}$$

Variabila $u = \frac{1}{n} \sum_{k=1}^{n} \left(\frac{x_k - \bar{x}}{\sigma}\right)^2$ este o sumă de n pătrate din care numai $(n - 1)$ sînt independente căci una din valorile x_i de pildă se poate calcula din relația $\bar{x} = \frac{1}{n} \sum_{k=1}^{n} x_k$ dacă se cunosc celelalte $(n - 1)$ valori $x_1, x_2 \ldots x_{i-1}, x_i, \ldots x_n$. În consecință, variabila u are o distribuție χ^2 cu $(n - 1)$ grade de libertate. Atunci $t = \left(\frac{\bar{x} - m}{\sigma}\right) : \sqrt{\frac{u}{n - 1}}$ are, după cum s-a văzut, o distribuție Student cu $(n - 1)$ grade de libertate. Densitatea sa :

$$f(t; n - 1) = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{(n-1)\pi}\,\Gamma\left(\frac{n-1}{2}\right)} \cdot \frac{1}{\left(1 + \frac{t^2}{n}\right)^{\frac{n}{2}}}, \quad -\infty < t < +\infty$$

nu conține nici unul din parametrii necunoscuți m sau σ. Graficul acestei densități este simetric față de axa ordonatelor și are înfățișarea de mai jos.

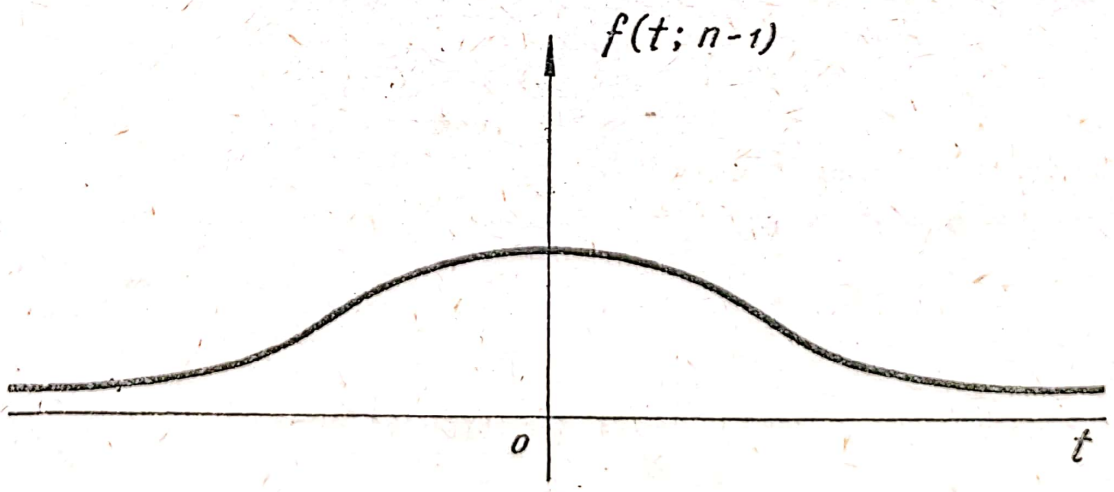

Fig. 50

În asemenea condiții, putem pune problema determinării unei valori $\bar{t}'$ a variabilei $\bar{t}$ astfel ca t să se găsească în intervalul $(-t', +\bar{t}')$ cu o probabilitate dată p, adică:

$$P(-t' < t < t') = \int_{-t'}^{t'} f(t; n-1)\, dt = p \qquad (1_6)$$

sau încă, ținînd seama de relația (1_5):

$$P\left[\bar{x} - t'\sqrt{\frac{\sum (x_k - \bar{x})^2}{n(n-1)}} < m < \bar{x} + t'\sqrt{\frac{\sum (x_k - \bar{x})^2}{n(n-1)}}\right] = p \qquad (1_7)$$

Valoarea t' se calculează cu ajutorul tabelelor care dau valorile distribuției Student. Cu ajutorul unei selecții de mărime n calculăm media $\bar{x}$ și obținem intervalul de confidență pentru m (cînd nu se cunoaște σ):

$$\left(\bar{x} - t'\sqrt{\frac{\sum (x_k - \bar{x})^2}{n(n-1)}},\ \bar{x} + t'\sqrt{\frac{\sum (x_k - \bar{x})^2}{n(n-1)}}\right) \qquad (1_7)$$

Să aplicăm cele expuse mai sus la următorul exemplu: Diametrele interioare ale unor inele de oțel produse de o anumită mașină au o distribuție normală. Pentru a estima media diametrului inelelor produse de mașină se face o selecție de 10 piese și se găsesc rezultatele de mai jos:

Tabloul 16

Proba	Diametrul în cm	Proba	Diametrul în cm
I	13,2	VI	16,1
II	16,4	VII	15,8
III	15,3	VIII	14,7
IV	14,2	IX	13,6
V	15,4	X	15,3

Să calculăm mai întîi numărul t' astfel ca, cu o probabilitate de 0,97, de pildă $\bar{t}$, să se găsească în intervalul $(-t', +t')$. Vom avea:

$$P(-t' < t < t') = F(t'; 9) - F(-t'; 9) = 0{,}97$$

unde F(t ; 9) este funcția de repartiție a distribuției Student cu 9 grade de libertate, adică :

$$F(t;\ 9) = \frac{\left(\frac{10}{2}\right)}{\sqrt{9\pi}\,\Gamma\left(\frac{9}{2}\right)} \int_{-\infty}^{t} \left(1 + \frac{x^2}{10}\right)^{-\frac{10}{2}} dx$$

În tabele se găsesc valorile F(t ; 9) pentru diferiți t. Întrucît între ordonatele punctelor $-t'$ și t' va trebui să se găsească 97% din aria de sub curba f(t ; 9), iar la stînga ordonatei lui $-t'$ și la dreapta ordonatei lui $+t'$, respectiv cîte 1,5% din această arie, rezultă că $\times t'$ se poate calcula din ecuația :

$$\int_{-\infty}^{+t'} f(t;\ 9)\ d\bar{t} + 0{,}015 = 1$$

deoarece aria întreagă este egală cu 1. Așadar, căutînd în tabela IV pe coloana 9 valoarea 0,985, găsim că pentru 0,984 corespunde valoarea $\bar{t}' \approx 2{,}6$.

Pentru calculul celorlalte elemente care dau intervalul de confidență, cu ajutorul datelor selecției formăm următorul tablou :

Tabloul 17

Proba	x_k	$x_k - \bar{x}$	$(x_k - \bar{x})^2$
I	13,2	− 1,8	3,24
II	16,4	+ 1,4	1,96
III	15,3	+ 0,3	0,09
IV	14,2	− 0,8	0,64
V	15,4	+ 0,4	0,16
VI	16,1	+ 1,1	1,21
VII	15,8	+ 0,8	0,64
VIII	14,7	− 0,3	0,09
IX	13,6	− 1,4	1,96
X	15,3	+ 0,3	0,09
	150		10,08

$\bar{x} = 150 : 10 = 15$

Scanned with OKEN Scanner

De aici deducem valorile :

$$\sum_{k=1}^{10} (x_k - \bar{x})^2 = 10{,}08$$

$$t' \sqrt{\frac{\sum (x_k - \bar{x})^2}{n\,(n-1)}} = 2{,}6 \sqrt{\frac{10{,}08}{10{,}19}} = 2{,}6 \times 0{,}35 = 0{,}91.$$

Obținem astfel intervalul de confidență 97% pentru media producției de piese fabricate de mașină, egal cu $15 \pm 0{,}91$, adică : $14{,}1 < m < 15{,}9$.

Aplicația II. Intervale de confidență pentre dispersia unei distribuții normale. Dacă $x_1, x_2, \ldots x_n$ sînt datele de selecție, variabila :

$$\chi^2 = \sum_{k=1}^{n} \frac{(x_k - \bar{x})^2}{\sigma} \qquad (1_8)$$

are o distribuție χ^2 cu $(n-1)$ grade de libertate.
Să determinăm două numere α și β astfel ca :

$$P(\alpha < \chi^2 < \beta) = p \qquad (1_9)$$

unde p este o probabilitate dată. Ținînd seama de (1_8), relația (1_9) se poate scrie :

$$P\left(\frac{\sum (x_k - \bar{x})^2}{\beta} < \sigma^2 < \frac{\sum (x_k - \bar{x})^2}{\alpha}\right) = p \qquad (1'_9)$$

Am obținut astfel intervalul de confidență cu probabilitatea p :

$$\left[\frac{\sum (x_k - \bar{x})^2}{\beta}, \frac{\sum (x_k - \bar{x})^2}{\alpha}\right] \qquad (1_{10})$$

Pentru determinarea numerelor α și β astfel ca lungimea intervalului (1_{10}) să fie minimă se folosesc tabelele distribuției χ^2 care dau probabilitățile :

$$P(\chi^2 > \chi_p^2) = \int_{\chi_p^2}^{\infty} f(\chi^2)\, d\chi^2 = p \qquad (1_{11})$$

pentru diferite valori ale lui p.
Așa de exemplu, dacă se caută intervale de confidență 97% se poate lua

$$\alpha = \chi^2_{0{,}985}, \quad \beta = \chi^2_{0{,}015}$$

așa încît, aria cuprinsă între ordonatele punctelor α și β să fie egală cu 97%, iar ariile situate la stînga ordonatei lui α și la dreapta ordonatei lui β să fie egale fiecare cu 1,5%.
Cu ajutorul tabelei III găsim pentru $n = 9$ valoarea $\alpha = x^2_{0,985} = 2{,}310$; pentru $x^2_{0,015}$ putem lua valoarea $x^2_{0,015} = 13{,}462 = \beta$. În exemplul precedent, pentru care am găsit $\sum (x_k - \bar{x})^2 = 10{,}08$ obținem intervalul de confidență 97%: $0{,}07 < \sigma^2 < 4{,}3$.

Aplicația III. Regiuni de confidență pentru valoarea medie și dispersia unei distribuții normale. Se știe că media de selecție și dispersia selecției sînt două variabile aleatorii independente, pentru populații normale. Așadar, variabilele $\bar{x}$ și $\sum_{k=1}^{n} (x_k - \bar{x})^2$ sînt independente în sensul calculului probabilităților.

Să presupunem, pentru exemplificare, că se caută o regiune A de confidență 97% pentru media m și dispersia σ^2 a populației normale Γ.

În acest scop, vom determina numărul α astfel ca:

$$P\left(-\alpha < \frac{\bar{x} - m}{\left(\frac{\sigma}{\sqrt{n}}\right)} < +\alpha\right) = \sqrt{0{,}97} \qquad (1_{12})$$

numerele β și γ astfel ca:

$$P\left(\beta < \frac{\sum (x_k - \bar{x})^2}{\sigma^2} < \gamma\right) = \sqrt{0{,}97} \qquad (1'_{12})$$

O dată determinate numerele α, β, γ, probabilitatea ca punctul aleator $M(m, \sigma^2)$ să se găsească în regiunea A, după regula de înmulțire a probabilităților evenimentelor independente, va fi dată de relația:

$$P\left\{-\alpha < \frac{\bar{x} - m}{\sigma/\sqrt{n}} < +\alpha;\ \ \beta < \frac{\sum (x_k - \bar{x})^2}{\sigma^2} < \gamma\right\} = 0{,}97 \qquad (1_{13})$$

Regiunea de confidență A se obține acum, construind față de sistemul de axe rectangulare m 0 σ^2 curbele:

$$\sigma^2 = \frac{n}{\alpha^2}(m - \bar{x})^2;\quad \sigma^2 = \frac{\sum (x_k - \bar{x})^2}{\beta};\quad \sigma^2 = \frac{\sum (x_k - \bar{x})^2}{\gamma} \qquad (1_{14})$$

Prima este o parabolă de gradul al doilea tangentă la axa 0 m cu vîrful în punctul de coordonate $(\bar{x}, 0)$ iar celelalte sînt două

Scanned with OKEN Scanner

drepte paralele cu axa absciselor 0 m. Aria din parabola limitată la aceste două parabole este regiunea de confidență 97% pe care o căutăm.

Observație. Această primă metodă de estimație a intervalelor sau regiunilor de confidență s-a bazat pe posibilitatea determinării unei funcții de datele selecției și care să nu depindă de parametrul ce trebuie estimat. (În exemplul tratat la început, abaterea normată $\bar{\eta}$ avea o distribuție independentă de parametrul de estimat m). O asemenea funcție nu există totdeauna; deci metoda nu poate fi aplicată decît într-un număr restrîns de cazuri. În paragraful următor vom expune o metodă generală pentru estimarea intervalelor și a regiunilor de confidență.

§ 3. Metoda generală

Să considerăm o populație oarecare Γ pe care s-a definit variabila aleatorie x cu densitatea $f(x; \lambda)$. O selecție $x_1, x_2, \ldots x_n$ ne va da indicatorul de probabilitate maximă $\lambda^* = \lambda^*(x_1, x_2 \ldots x_n)$, în funcție de datele selecției. Acest indicator am văzut cum se calculează: se determină densitatea asociată datelor de selecție P_n și se anulează derivata $\frac{\partial P_n}{\partial \lambda}$ etc. Am văzut că λ^* depinde și de parametrul λ al distribuției. Să notăm densitatea λ^* prin $\varphi(\lambda^*; \lambda)$.

Pentru a găsi un anumit interval de confidență, 97% de pildă vor trebui determinate două funcții:

$$\alpha_1 = \alpha_1(\lambda), \quad \alpha_2 = \alpha_2(\lambda) \tag{2}$$

în așa fel ca:

$$P(\alpha_1 < \lambda^* < \lambda_2) = \int_{\alpha_1}^{\alpha_2} \varphi(\lambda^*; \lambda)\, d\lambda^* = 0{,}97 \tag{2_1}$$

Geometric, aceasta însemnează că ordonatele acestor puncte detașează 97% din aria mărginită de curba $\varphi(\lambda^*; \lambda)$; restul de 3% din arie îl descompunem în două părți echivalente, așa încît α_1 și α_2 verifică ecuațiile:

$$P(\lambda^* < \alpha_1) = \int_{-\infty}^{\alpha_1} \varphi(\lambda^*; \lambda)\, d\lambda^* = 0{,}015$$

$$P(\lambda^* > \alpha_2) = \int_{\alpha_2}^{+\infty} \varphi(\lambda^*; \lambda)\, d\lambda^* = 0{,}015 \tag{2_2}$$

Scanned with OKEN Scanner

Aşadar, vom determina funcţiile $\alpha_1(\lambda)$ şi $\alpha_2(\lambda)$ din aceste ecuaţii; pe λ^* îl determinăm cu ajutorul unei selecţii, fie $\lambda^{*\prime}$ valoarea găsită.

Ecuaţiile:

$$\lambda^{*\prime} = \alpha_1(\lambda), \quad \lambda^{*\prime} = \alpha_2(\lambda) \qquad (2_3)$$

vor da intervalul de confidenţă 97%, căutat. Dacă λ_1 este valoarea lui λ din prima ecuaţie şi λ_2 din cea de a doua ecuaţie (2_3) intervalul de confidenţă este (λ_1, λ_2). Se poate ca sistemul (2_3) să dea mai multe intervale; în asemenea împrejurări, ţinînd seama şi de alte informaţii suplimentare obţinute din selecţie, este posibil să se aleagă un interval de confidenţă unic.

Rezultatele stabilite pînă acum devin mai intuitive folosind o interpretare geometrică a lor. În acest scop, alegem λ ca variabilă şi λ^* funcţie. Faţă de un sistem de axe rectangulare $\lambda\lambda^*$, funcţiile $\lambda^* = \alpha_1(\alpha)$ şi $\lambda^* = \alpha_2(\lambda)$ se reprezintă ca în figura alăturată

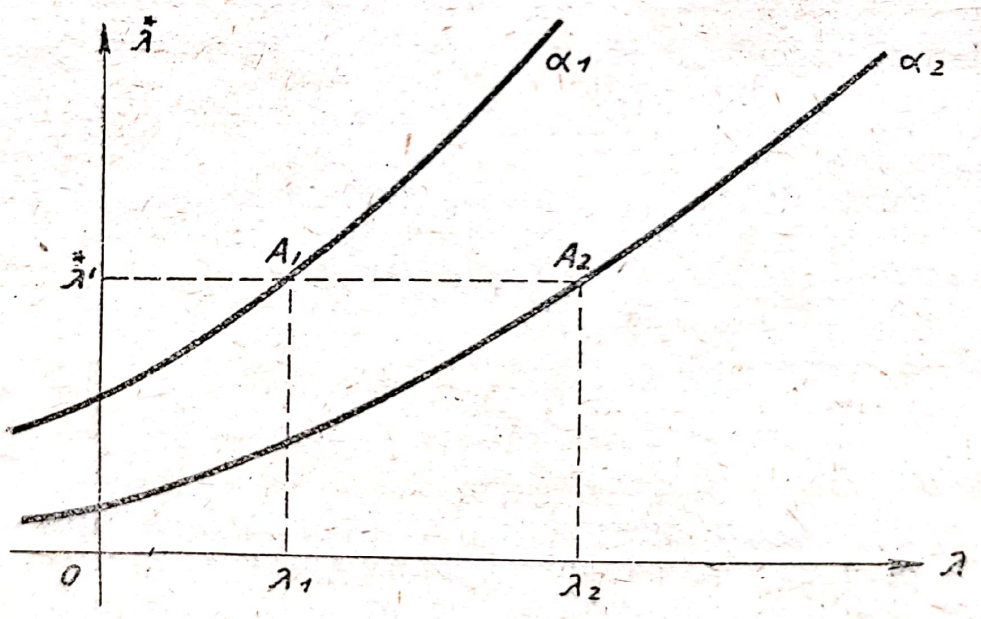

Fig. 51

Paralela $\lambda^* = \lambda^{*\prime}$ (cunoscută din datele selecţiei) va tăia aceste curbe în două puncte A_1 şi A_2 ale căror abscise $\lambda_1 = 0A_1$ şi $\lambda_2 = 0A_2$ sînt limitele intervalului de confidenţă 97%.

Exemplu : Diametrul x al unei piese fabricată de o anumită mașină are distribuția :

$$f(x;\ \lambda) = \frac{2(\lambda - x)}{\lambda^2},\quad 0 < \chi = \lambda \tag{3}$$

Să estimăm parametrul λ printr-o selecție formată dintr-un singur element. Urmînd metoda indicată vom calcula mai întîi indicatorul $\lambda^*(x_1)$, din ecuația :

$$\frac{\partial f(x_1;\lambda)}{\partial \lambda} = \frac{2}{\lambda^3}(2x_1 - \lambda) = 0 \tag{3_1}$$

De unde deducem $\lambda^* = 2x_1$, (x_1 fiind dat de selecție). Să determinăm acum distribuția $\varphi(\lambda^*;\lambda)$ a lui λ^*.
Avem formula cunoscută :

$$\varphi(\lambda^*;\ \lambda) = f[x_1(\lambda^*);\ \lambda]\ \frac{dx_1}{d\lambda^*} \tag{3_2}$$

și acum :

$$f[x_1(\lambda^*);\ \lambda] = \frac{2\left(\lambda - \frac{\lambda^*}{2}\right)}{\lambda_2} = \frac{2\lambda - \lambda^*}{\lambda^2}$$

$$x_1 = \frac{\lambda^*}{2},\ \text{deci}\quad \left|\frac{dx_1}{d\lambda^*}\right| = \frac{1}{2}$$

se găsește, înlocuind în relația (3_2), distribuția :

$$\varphi(\lambda^*;\ \lambda) = \frac{2\lambda - \lambda^*}{2\lambda^2} \tag{3_3}$$

În această distribuție λ^* ia valori de la 0 la 2 λ, deoarece $x_1 = \frac{\lambda^*}{2}$ variază în relația (3) de la 0 la λ.

Dacă am dori de pildă să obținem intervale de confidență 95%, vom determina pe $\alpha_1(\lambda)$ și $\alpha_2(\lambda)$ din ecuațiile :

$$\int_0^{\alpha_1} \varphi(\lambda^*;\ \lambda)\ d\lambda^* = 0{,}025 \tag{3_4}$$

$$\int_{\alpha_2}^{2\lambda} \varphi(\lambda^*;\ \lambda)\ d\lambda^*) = 0{,}025$$

Scanned with OKEN Scanner

Efectuînd prima integrală găsim :

$$\int_0^{\alpha_1} \frac{2\lambda - \lambda^*}{2\lambda^2} d\lambda^* = \frac{\alpha_1}{\lambda} - \frac{\alpha_1^2}{4\lambda^2} = 0{,}025$$

De unde ecuația în α_1 : $\alpha_1^2 - 4\lambda\alpha_1 + 4\lambda^2 \cdot 0{,}025 = 0$ cu rădăcinile $\alpha_1 = 2\lambda \pm 2\lambda\sqrt{0{,}975}$.

În mod analog, se găsește $\alpha_2 = 2\lambda \pm 2\lambda\sqrt{0{,}025}$.
Deoarece $0 < \lambda^* < 2\lambda$ vom lua rădăcinile :

$$\alpha_1 = 2(1 - \sqrt{0{,}975})\,\lambda,\ \alpha_1 = 2(1 - \sqrt{0{,}025})\,\lambda \qquad (3_5)$$

În acest caz, funcțiile $\lambda^* = \alpha_1(\lambda), \lambda^* = \alpha_2(\lambda)$ reprezintă două drepte ce trec prin origine. Să presupunem că o selecție a dat valoarea $x_1 = 1{,}5$ atunci :

$$\lambda^* = 2x_1 = 3.$$

Intersectînd dreapta $\lambda^* = 3$ cu dreptele (3_5) sd găsește intervalul de confidență 95%, $\left[\frac{3}{2(1-\sqrt{0{,}025})}, \frac{3}{2(1-\sqrt{0{,}975})}\right]$.

Probabilitatea ca parametrul adevărat λ să se găsească în acest interval este egală cu 0,95.

§ 4. Selecții de volum mare

Se știe că în acest caz, indicatorul λ^* are distribuție normală, în condiții foarte generale, cu valoarea medie

$$M(\lambda^*) = \lambda \qquad (4)$$

și dispersia :

$$D^2(\lambda^*) = -\frac{1}{n\,M\left[\frac{\partial^2 \ln f(x;\lambda)}{\partial \lambda^2}\right]} \qquad (4_1)$$

Fie p o probabilitate dată (0,97 de pildă); probabilitatea ca abaterea normată

$$\frac{\lambda^* - M(\lambda^*)}{D(\lambda)^*} = \frac{\lambda^* - \lambda}{D(\lambda^*)} \qquad (4_2)$$

să ia o valoare cuprinsă în intervalul $(-a, +a)$ este aproximativ egală cu integrala.

$$\frac{1}{\sqrt{2\pi}} \int_{-a}^{+a} e^{-\frac{1}{2}t^2} dt = 2\,\Phi(a)$$

unde $\Phi(a)$ este funcția cunoscută a lui Laplace. Punînd condiția ca această probabilitate să fie egală cu p, determinăm pe a în așa fel ca:

$$P\left(-a < \frac{\lambda^* - \lambda}{D(\lambda^*)} < +a\right) = p \qquad (4_4)$$

și în consecință obținem intervalul

$$P\left(\lambda^* - a\,D(\lambda^*) < \lambda < \lambda^* + a\,D(\lambda^*)\right) = p$$

Așa de exemplu pentru o distribuție normală

$$f(x;\lambda) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\lambda}{\sigma}\right)^2}, \quad -\infty < x < +\infty$$

în care este dat σ, indicatorul $\lambda^* = \bar{x}$ media de selecție, Cum $\bar{x}$ are o distribuție normală cu media λ și dispersia $\frac{\sigma^2}{n}$ avem distribuția lui λ^*:

$$\frac{\sqrt{n}}{\sqrt{2\pi}\,\sigma} e^{-\frac{n}{2}\left(\frac{\lambda^* - \lambda}{\sigma}\right)^2}$$

De unde deducem probabilitatea:

$$P\left(\lambda^* - \frac{a\,\sigma}{\sqrt{n}} < \lambda < \lambda^* + \frac{a\,\sigma}{\sqrt{n}}\right) = \int_{-a}^{+a} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}$$

Scanned with OKEN Scanner

Dacă de pildă $p = 0{,}95$ atunci $2\Phi(a) = 0{,}95$ de unde rezultă din tabela valorilor funcției $\Phi(a)$ valoarea $a = 1{,}96$. Pentru $n = 100$ avem deci: $P(\bar{x} - 0{,}196\,\sigma < \lambda < \bar{x} + 0{,}196) = 0{,}85$ Așadar cu o porbabilitate de 95% parametrul λ se va găsi în intervalul $(\bar{x} - 0{,}2\sigma,\ x + 0{,}2\sigma)$ în jurul mediei de selecție.

Metoda pe care am expus-o în acest paragraf se poate folosi și pentru distribuțiile cu mai mulți parametri. Să considerăm de pildă o distribuție cu doi parametri λ_1 și λ_2. În planul λ_1^* λ_2^* putem determina un domeniu Δ așa fel ca:

$$P(\lambda_1^*,\ \lambda_2^* \varepsilon \Delta) = \int\int_{\Delta} \varphi(\lambda_1^*,\ \lambda_2^*;\ \lambda_1,\ \lambda_2)\ d\lambda_1^*\ d\lambda_2^* \qquad (5)$$

Mulțimea tuturor perechilor de valori posibile ale parametrilor λ_1 și λ_2 dă naștere unui domeniu cu patru dimensiuni λ_1, λ_2, λ_1^*, λ_2^*. Dacă se face o selecție și se calculează $\lambda_1^{*'}$ $\lambda_2^{*'}$ intersecția hiperplanelor $\lambda_1^* = \lambda_1^{*'}$, $\lambda_2^* = \lambda_2^{*'}$ cu Δ va determina un domeniu bidimensional care proiectat de planul λ_1 λ_2 dă o regiune de confidență pentru parametrii λ_1 și λ_2.

Observație. Metode de estimarea parametrilor și intervalelor sau regiunilor de confidență sînt relativ numeroase. În noțiunile cu caracter elementar pe care le-am dat ne-am folosit îndeosebi de indicatorii de probabilitate maximă. Acești indicatori se bucură de proprietatea importantă că pentru selecții de volum mare au dispersia minimă și în consecință intervalele sau regiunile de confidență determinate cu ajutorul lor vor fi în general mai mici decît atunci cînd sînt calculate pe baza altor indicatori (de pildă, anumite funcții de indicatori de probabilitate maximă).

Probleme

1. Să se determine un interval de confidență 95% pentru valoarea medie a unei distribuții normale cu $\sigma = 5$ pe baza selecției de patru elemente: (1,5; 2,1; 1,9; 1,7)

2. În problema de mai sus să se determine intrevalul de confidență dacă σ este necunoscut.

3. Știind că rezistența în grame a unui fir cu grosimea $\frac{7}{39}$ cm pe baza a pentru selecții a fost stabilită la 63, 71, 52, 49 să se estimeze rezistența medie cu un interval de confidență 97%, presupunînd distribuția normală.

4. În problema precedentă să se estimeze punctul în care se rup în medie numai 3% din cele patru selecții.

Scanned with OKEN Scanner

5. Cu datele din problema 3 să se estimeze σ și σ^2 printr-un interval de confidență 95%.

6. Distribuția abaterilor dimensiunilor diametrelor unei piese de la medie este normală. În mici selecții luate din oră în oră s-au stabilit mediile :
31,3 cm, 27,9 cm, 35,2 cm, 34 cm. 26cm.

Să se determine o regiune de confidență 90% pentru estimarea simultană a lui m și σ^2.

Scanned with OKEN Scanner

TABELA I

Valorile funcției $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3043	3011	2989	2966	2943	2920
0,8	2897	3874	3850	3827	3803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	2761	1256	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0399	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	1107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Scanned with OKEN Scanner

TABELA II

Valorile funcției $2\Phi(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-\frac{1}{2}x^2} dx$

	0	1	2	3	4	5	6	7	8	9
0,0	0,0 0000	0798	1596	2393	3191	3988	4784	5581	6376	7171
0,1	7966	8759	9552	0343	1134	1924	2712	3499	4285	5069
0,2	0,1 5852	6633	7413	8191	8967	9741	0514	1284	2052	2818
0,3	0,2 3582	4344	5103	5860	6614	7366	8115	8862	9605	0346
0,4	0,3 1084	1819	2552	3280	4006	4729	6448	6164	6877	7587
0,5	8292	8995	9694	0389	1080	1768	2452	3132	3809	4481
0,6	0,4 5149	5814	6474	7131	7783	8431	9075	9714	0350	0981
0,7	0,5 1607	2230	2848	3461	4070	4675	5275	5870	6161	7047
0,8	7629	8206	8778	9346	9909	0468	1021	1570	2114	2653
0,9	0,6 3188	3718	4243	4763	5278	5789	6294	6795	7291	7783
1,0	8269	8750	9227	9699	0166	0628	1086	1538	1986	2429
1,1	0,7 2867	3300	3729	4152	4571	4986	5395	5800	6200	6595
1,2	6986	7372	7754	8130	8502	8870	9233	9592	9945	0295
1,3	0,8 0640	0980	1316	1648	1975	2298	2674	2931	3241	3547
1,4	3849	4146	4439	4728	5013	5294	5571	5844	6113	6378
1,5	6639	6696	7149	7398	7644	7886	8124	8358	8589	8817
1,6	9040	9260	9477	9690	9899	0106	0309	0508	0704	0957
1,7	0,9 1087	1273	1457	1637	1814	1988	2159	2327	2492	2655
1,8	2814	2970	3124	3275	3423	3569	3711	3852	3989	4124
1,9	4257	4387	4514	4639	4762	4882	5000	5116	5230	5341
2,0	5450	5557	5662	5764	5865	5964	6060	6155	6247	6338
2,1	6427	6514	6599	6683	6795	6844	6923	6999	7074	7148
2,2	7219	7289	7358	7425	7491	7555	7618	7679	7739	7798
2,3	7855	7911	7966	8019	8072	8123	8172	8221	8269	8315
2,4	8360	8405	8448	8490	8531	8571	8611	8649	8686	8723
2,5	8758	8793	8826	8859	8891	8923	8953	8983	9012	9040
2,6	9068	9095	9121	9146	9171	9195	9219	9241	9263	9285
2,7	9307	9327	9347	9367	9386	9404	9422	9439	9496	9473
2,8	9489	9505	9520	9535	9549	9563	9576	9590	9602	9615
2,9	9627	9639	9650	9661	9672	9682	9692	9702	9712	9721
3,0	9730	9739	9747	9755	9763	9771	9779	9786	9793	9800
3,1	9806	9813	9819	9825	9831	9837	9842	9848	9853	9858
3,2	9863	9867	9872	9876	9880	9885	9889	9892	9896	9900
3,3	9903	9907	9910	9913	9916	9919	9922	9925	9928	9930
3,4	9933	9935	9937	9940	9942	9944	9946	9948	9950	9952
3,5	9953	9955	9957	9958	9960	9961	9963	996	9966	9967
3,6	9968	9969	9971	9972	9973	9974	9975	9976	9977	9978
3,7	9978	9979	9980	9981	9982	9982	9983	9984	9984	9985
3,8	9986	9986	9987	9987	9988	9988	9989	9989	9000	9990
3,9	9990	9991	9992	9992	9992	9992	9992	9993	9993	9993

Scanned with OKEN Scanner

Valorile probabilităților

n	p = 99	98	95	90	80	70	50
1	0,000	0,001	0,004	0,016	0,064	0,148	0,455
2	0,020	0,040	0,103	0,211	0,446	0,713	1,386
3	0,115	0,185	0,352	0,548	1,005	1,424	2,366
4	0,297	0,429	0,711	1,064	1,649	2,195	3,357
5	0,554	0,752	1,145	1,610	2,343	3,000	4,531
6	0,872	1,134	1,635	2,204	3,070	3,828	5,348
7	1,239	1,564	2,167	2,833	3,822	4,671	6,346
8	1,646	2,032	2,733	3,490	4,594	5,527	7,344
9	2,088	2,532	3,325	4,168	5,380	6,393	8,343
10	2,558	3,059	3,940	4,865	6,179	7,267	9,342
11	3,053	3,609	4,575	5,578	6,989	8,148	10,341
12	3,571	4,179	5,226	6,304	7,807	9,034	11,340
13	4,107	4,765	5,892	7,042	8,634	9,926	12,340
14	4,660	5,368	6,571	7,790	9,467	10,821	13,339
15	5,229	5,985	7,261	8,547	10,307	11,721	14,339
16	5,812	6,614	7,962	9,312	11,152	12,624	13,338
17	6,408	7,255	8,672	10,085	12,002	13,531	16,336
18	7,015	7,906	9,390	10,865	12,857	14,440	17,338
19	7,633	8,567	10,117	11,651	13,716	15,352	18,338
20	8 260	9,237	10,851	12,443	14,578	16,266	19,337
21	8,897	9,915	11,591	13,240	15,445	17,182	20,337
22	9,542	10,600	12,338	14,041	16,314	18,101	21,337
23	10,196	11,293	13,091	14,848	17,187	19,021	22,337
24	10,856	11,992	13,848	15,659	18,062	19,943	23,337
25	11,524	12,697	14,611	16,473	18,940	20,867	24,337
26	12,198	13,409	15,379	17,292	19,820	21,792	25,336
27	12,879	14,125	16,151	18,114	20,703	22,719	26,336
28	13,565	14,847	16,928	18,939	21,588	23,647	27,336
29	14,256	15,574	17,708	19,768	22,475	24,577	28,336
30	14,953	16,306	18,493	20,599	23,364	25,508	29,336

Scanned with OKEN Scanner

LA III

$$P(\chi^2 > \chi_p^2) = \int_{\chi_p^2}^{\infty} f(\chi^2)\, d\chi^2$$

30	20	10	5	2	1	0,1
1,074	1,642	2,706	3,841	5,412	6,635	10,827
2,408	3,219	4,605	5,991	7,842	9,210	13,815
3,605	4,642	6,251	7,815	9,837	11,341	16,268
4,878	5,989	7,779	9,488	11,668	13,277	18,645
6,064	7,289	9,236	11,070	13,388	15,086	20,517
7,231	8,558	10,645	12,591	15,033	16,812	22,457
8,383	9,903	12,017	14,067	16,622	18,475	24,322
9,524	11,030	13,362	15,507	18,168	20,000	26,125
10,656	12,242	14,684	16,919	19,679	21,666	27,877
11,781	13,442	15,987	18,307	21,162	23,209	29,588
12,899	14,631	17,275	19,675	22,618	24,725	31,264
14,011	15,812	18,549	21,026	24,054	26,217	32,909
15,119	16,985	19,812	22,362	25,472	27,688	34,528
16,222	18,151	21,064	23,685	26,873	29,141	36,123
17,322	19,311	22,307	24,996	28,259	30,578	37,697
18,418	20,465	23,542	26,296	29,633	32,000	39,252
19,511	21,615	24,769	27,587	30,995	33,409	40,790
20,601	22,760	25,989	28,869	32,346	34,805	42,312
21,689	23,900	27,204	30,144	33,687	36,191	43,820
22,775	25,038	28,412	31,410	35,020	37,566	45,315
23,858	26,171	29,615	32,671	36,343	38,932	46,797
24,939	27,301	30,813	33,924	37,659	40,289	48,268
26,018	28,429	32,007	35,172	38,968	41,638	49,728
27,096	29,553	33,196	36,415	40,270	42,980	51,179
28,172	30,675	34,382	37,652	41,566	44,314	52,620
29,246	31,795	35,563	38,885	42,856	45,642	54,052
30,319	32,912	36,741	40,113	44,140	46,963	55,476
31,319	34,027	37,916	41,337	45,419	48,278	56,893
32,461	35,139	39,087	42,557	46,693	49,588	58 302
33,530	36,250	40,256	43,773	47,962	50,892	59,703

Scanned with OKEN Scanner

TABELA IV

Valorile funcției de repartiție pentru distribuția t cu (n — 1) grade de libertate:

$$F(t; n-1) = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{(n-1)\pi}\,\Gamma\left(\frac{n-1}{2}\right)} \int_{-\infty}^{t} \left(1 + \frac{x^2}{n-1}\right)^{-\frac{n}{2}} dx$$

x \ n	2	3	4	5	6	7	8	9	10	11
0,0	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500
0,1	532	535	537	537	538	538	538	539	539	539
0,2	563	570	573	574	575	576	576	577	577	577
0,3	593	604	608	610	612	613	614	614	614	615
0,4	621	636	642	645	647	648	650	650	651	651
0,5	648	667	674	678	681	683	684	685	686	686
0,6	672	695	705	710	713	715	716	717	718	719
0,7	694	722	733	739	742	745	747	748	749	750
0,8	715	746	759	766	770	773	775	777	778	779
0,9	733	768	783	790	795	799	801	803	804	805
1,0	750	789	804	813	818	822	825	827	828	830
1,1	765	807	824	834	839	843	846	848	850	851
1,2	779	824	842	852	858	862	865	868	870	871
1,3	791	838	858	868	875	879	883	885	887	889
1,4	803	852	872	883	890	894	898	900	902	904
1,5	813	864	885	896	903	908	911	914	916	918
1,6	822	875	896	908	915	920	923	926	928	930
1,7	831	884	906	918	925	930	934	936	938	940
1,8	839	893	915	927	934	939	943	945	947	949
1,9	846	901	923	935	942	947	950	953	955	957

Scanned with OKEN Scanner

(Continuare tabela IV)

x \ n	2	3	4	5	6	7	8	9	10	11
2,0	852	908	930	942	949	954	957	960	962	963
2,2	864	921	942	954	960	965	968	970	972	974
2,4	874	931	952	963	969	973	976	978	980	981
2,6	883	938	960	970	976	980	982	984	986	987
2,8	891	946	966	976	981	984	987	988	990	991
3,0	898	952	971	980	985	988	990	992	992	993
3,2	904	957	975	984	988	991	992	994	995	995
3,4	909	962	979	986	990	993	994	995	996	997
3,6	914	965	982	989	992	994	996	996	997	998
3,8	918	969	984	990	994	996	997	997	998	998
4,0	922	971	986	992	995	996	997	998	998	999
4,2	926	974	988	993	996	997	998	998	999	999
4,4	929	976	989	994	996	998	998	999	999	999
4,6	932	978	990	995	997	998	999	999	999	1,000
4,8	935	980	991	996	998	998	999	999	1,000	
5,0	937	981	992	996	998	999	999	1,000		
5,2	940	982	993	997	998	999	999			
5,4	942	984	994	997	998	999	1,000			
5,6	944	985	994	998	999	999				
5,8	946	986	995	998	999	999				
6,0	947	987	995	998	999	1,000				

Scanned with OKEN Scanner

Scanned with OKEN Scanner

TABLA DE MATERII

DIN PARTEA AUTORULUI 3

INTRODUCERE 5

CAPITOLUL I **Noţiuni introductive** 9

§ 1. Problema selecţiei 9
§ 2. Serii statistice. Reprezentări geometrice 11
§ 3. Momentele empirice ale distribuţiilor statistice 23
§ 4. Medii statistice. Calculul simplificat 26
Probleme 40

CAPITOLUL II **Noţiuni de bază în teoria probabilităţilor** 42

CAPITOLUL III **Probabilităţi** 52

§ 1. Definiţia axiomatică a probabilităţii 52
§ 2. Definiţia clasică a probabilităţii 54
§ 3. Probabilităţi condiţionate 56
§ 4. Evenimente compatibile 63
Probleme 65

CAPITOLUL IV **Distribuţii discrete** 68

§ 1. Variabile aleatorii 68
§ 2. Distribuţii unidimensionale. 72
§ 3. Distribuţii multidimensionale. 81
Probleme 91

CAPITOLUL V **Distribuţii continue** 93

§ 1. Distribuţii unidimensionale 93
§ 2. Distribuţii continue multidimensionale 110
Probleme 123

CAPITOLUL VI **Valori medii. Momente** 125

§ 1. Definiţia valorii probabile pentru variabilele aleatorii de tip discret. 125
§ 2. Funcţii monotone de variabile aleatorii 132
§ 3. Momentele distribuţiilor unidimensionale 137
§ 4. Funcţia generatoare de momente 156
§ 5. Momentele distribuţiilor multidimensionale. 162
§ 6. Metoda funcţiei caracteristice. 166

§ 7. — Problema momentelor 170
Probleme 173

CAPITOLUL VII **Distribuţia normală multidimensională şi distribuţiile funcţiilor de variabile aleatorii** 175
§ 1. Noţiuni sumare asupra formelor lineare şi patratice 175
§ 2. Distribuţia normală bidimensională. 183
§ 3. Distribuţii de funcţii de variabile aleatorii 201
Probleme 211

CAPITOLUL VIII **Legea numerelor mari. Teoreme de convergenţă**. . . . 214

§ 1. Convergenţa în probabilitate 214
§ 2. Teoreme de convergenţă. 224
Probleme. 131

CAPITOLUL IX **Selecţii aleatorii** 232

§ 1. Distribuţii de selecţie 232
§ 2. Momentele selecţiei 234
§ 3. Selecţii de volum redus (selecţii mici). 246
Probleme 255

CAPITOLUL X **Teoria estimaţiei** 257

§ 1. Probleme parametrice 257
§ 2. Estimaţii eficiente 257
§ 3. Estimaţii suficiente 263
§ 4. Metoda momentelor 266
§ 5. Metoda verosimilităţii maxime. 267
§ 6. Distribuţia indicatorilor de probabilitate maximă 272
§ 7. Estimaţii pentru selecţiile nerepetate 281

CAPITOLUL XI **Estimarea intervalelor** 287

§ 1. Intervale de confidenţă 287
§ 2. Prima metodă (elementară) pentru determinarea intervalelor de confidenţă. 287
§ 3. Metoda generală 297
Probleme, ... 302

Redactor de carte: V. Matei
Tehnoredactor: Gh. Popovici

Dat la cules 28.03.1957. Bun de tipar 22.07.1957. Tiraj 2000+110. Hîrtie semivelină de 80 gr./m.p. Ft. 61×86/16. Coli editoriale 11.99. Coli de tipar 19,50 A.: 0823/1957. Pentru bibliotecile mari indicele de clasificare 31:51. Pentru bibliotecile mici: 31:51.

Tiparul executat la Întreprinderea poligrafică nr. 3, B-dul 6 Martie nr. 29—Bucureşti R.P.R.

Scanned with OKEN Scanner

E R A T Ă

Pag.:	Rîndul:	În loc de:	Se va citi:
24	1 de jos	$(\alpha_h - \overline{x})^2$	$(x_h - \overline{x})^2$
29	col. 2 tablou 14	$\alpha_h = d_h$	$\xi_h = \alpha_h$
34	7 de jos	$m = {}^n x_1 x_2 \ldots x_n$	$m = \sqrt[n]{x_1 x_2 \ldots x_n}$
44	1 de jos	$P \varepsilon K$	$P \notin K$
78	12 de jos	$P (M \varepsilon [a, x])$	$P (M \varepsilon [a, x))$
103	1 de jos	$(m + a) =$	$f (m + a) =$
103	2 de sus	$\frac{x - m}{\sigma} = \sigma$	$\frac{x - m}{\sigma} = y$
143	12 și 13 de jos	ϑ_r	ν_r
144	11 de sus	$\frac{a^2}{\rho^2}$	$\frac{a^2}{\sigma^2}$
160	11 de sus	$\frac{t}{\rho}$	$\frac{t}{\sigma}$
229	9 de jos	$\frac{1}{6} \Phi$	$\frac{1}{6} \theta$
229	5 de jos	$\log (1 + \alpha)$	$\ln (1 + \alpha)$
240	4 de jos	$\sum \left\| \frac{x_h}{n} - \frac{m}{n} \right\|$	$\sum M \left\| \frac{x_h}{n} - \frac{m}{n} \right\|$
241 / 242	11 de jos și 3 de jos	$n^{\frac{1}{2}}$	$\frac{1}{n^2}$
243	10 de sus	$m_4 + \frac{3(n-1)(2n-3)}{n^3} m_2^3$	$\mu_4 + \frac{3(n-1)(2n-3)}{n^3} \mu_2^3$
255	5 de jos	$\Gamma\left(\frac{n}{2}\right)^{\frac{n}{\sigma}}$	$\Gamma\left(\frac{n}{2}\right) \sigma^n$
259	10 de jos	$\int_{-\infty}^{+\infty} \frac{\ln f (x ; \lambda)}{\partial \lambda}$	$\int_{-\infty}^{+\infty} \frac{\partial \ln f (x ; \lambda)}{\partial \lambda}$
261	8 de sus	$f (x ; \lambda) = \sigma^{-\lambda} \cdot \frac{\lambda^*}{x!}$	$f (x ; \lambda) = e^{-\lambda} \frac{\lambda^x}{x!}$
262	9 de jos	$\frac{2 \ln f (x ; p)}{p^2}$	$\frac{\partial^2 \ln f (x ; p)}{\partial p^2}$

Scanned with OKEN Scanner